# Incertitudes expérimentales

F. Levrier

**Résumé**

On présente dans ce texte les outils de base permettant d'apprécier de manière quantitative les résultats de manipulations expérimentales telles que celles abordées à l'agrégation. Après avoir défini la notion d'incertitude et rappelé quelques éléments de probabilités et de statistiques, on montre comment quantifier les incertitudes associées au caractère aléatoire des processus de mesure à l'aide d'une étude statistique, et celles associées à une grandeur physique obtenue par calcul à partir d'un certain nombre d'autres grandeurs mesurées directement. On donne ensuite plusieurs exemples et remarques utiles, notamment l'évaluation des incertitudes à partir des indications des appareils de mesure. On aborde enfin les principes de la modélisation et de la vérification d'une loi physique par l'ajustement de données expérimentales. Autant que possible, on illustre ces différentes parties par des exemples concrets rencontrés en TP.

*Ce texte se fonde en grande partie sur celui rédigé pour le même cours par François-Xavier Bally et Jean-Marc Berroir, également publié dans le BUP 928, ainsi que sur le cours "Bruits et signaux" de Didier Pelat au Master "Astronomie, Astrophysique et Ingénierie Spatiale". Il fait également référence aux documents disponibles sur eduscol pour l'enseignement de ces notions au lycée.*

## Table des matières

# 1 Erreurs et incertitudes

La notion d'incertitude est essentielle dans la démarche expérimentale. Sans elle, on ne peut juger de la qualité d'une mesure, de sa pertinence ou de sa compatibilité avec une loi physique. C'est pourquoi il est indispensable, notamment dans le cadre de l'épreuve de montage à l'agrégation de physique, d'évaluer correctement et de discuter les incertitudes affectant les résultats des mesures présentées.

Pour poser les bases, considérons une grandeur physique $X$ dont on souhaite connaître la valeur vraie [1], notée $x_0$, au cours d'une expérience. On ne connaît bien entendu pas cette valeur

1. L'usage de la notion de "valeur vraie" est maintenant déconseillé, en raison du fait que toute grandeur fluctue naturellement dans le temps. On privilégiera la notion de "valeur de référence".

puisqu'on la cherche. À l'aide d'un appareil adéquat [2], on mesure une valeur $x$, a priori différente de $x_0$. Il convient tout d'abord de distinguer *l'erreur* $\epsilon = x - x_0$, qui est la différence entre la valeur mesurée et la valeur vraie [3], de *l'incertitude* $\delta x$, qui est une notion statistique dont on précisera la définition plus bas [4]. Beaucoup de scientifiques confondent ces deux termes et parlent à tort de calculs d'erreurs au lieu de calculs d'incertitudes.

## 1.1 Erreurs aléatoires

Supposons que la mesure de $X$ puisse être faite à plusieurs reprises de manière indépendante. Pour fixer les idées, prenons le cas de la mesure de la période $T$ des oscillations d'un pendule en opérant avec un chronomètre manuel. On constate qu'en répétant les mesures on trouve des résultats légèrement différents à chaque fois, dûs surtout aux retards ou aux avances de déclenchement et d'arrêt du chronomètre par l'expérimentateur [5]. Ces écarts vont modifier la valeur mesurée de $T$. Ce phénomène sera détecté par une étude statistique : si l'on fait un grand nombre de mesures dans des conditions identiques [6], les résultats $t_1$, $t_2$, ..., $t_n$ de ces mesures vont se répartir suivant une distribution de probabilité (voir la section 2). On parle dans ce cas d'*erreur aléatoire*. Dans le cas où les erreurs sont purement aléatoires, la distribution de probabilité des résultats de mesure est répartie autour de la valeur vraie $t_0$. Dans la table 1, on donne un exemple d'une série de mesures de $T$ entachées d'erreurs purement aléatoires, alors que la valeur vraie est $t_0 = 2\,\mathrm{s}$. On constate que les mesures se répartissent bien aléatoirement autour de $t_0$.

La visualisation de la répartition des résultats se fait au moyen d'*histogrammes*, représentation en colonnes jointives avec en abscisse une échelle des valeurs représentées et en ordonnée le nombre de valeurs (ou leur fréquence d'apparition) se situant dans un intervalle donné. On peut y attacher quelques définitions [6] :

- *étendue des observations* : la valeur maximale moins la valeur minimale observée
- *classe* $i$ : un des intervalles au sein desquels on compte le nombre d'observations
- *étendue* $w_i$ *d'une classe* $i$ : la largeur de l'intervalle correspondant (on prend souvent la même pour toutes les classes)
- *effectif* $n_i$ : le nombre d'observations qui se trouve dans la classe $i$
- *fréquence* $f_i = n_i/N$ : la proportion des $N$ observations qui se trouve dans la classe $i$.

TABLE 1 – Série de mesures d'une période d'oscillation d'un pendule de période $t_0 = 2\,\mathrm{s}$, effectuées avec un chronomètre manuel précis au centième de seconde, sans erreur systématique. La moyenne des mesures est $\langle t_i \rangle = 2,03\,\mathrm{s}$.

| $i$ | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 |
|---|---|---|---|---|---|---|---|---|---|---|
| $t_i$ (s) | 2,12 | 1,88 | 1,98 | 1,95 | 1,92 | 2,06 | 2,08 | 2,16 | 2,03 | 2,11 |

2. Notons qu'il faut utiliser autant que possible les appareils *dédiés à la mesure*. On privilégiera par exemple la mesure d'une tension au voltmètre plutôt qu'en utilisant les curseurs d'un oscilloscope : comme leurs noms l'indiquent, l'un sert à mesurer, l'autre à visualiser...

3. On ne connaît pas plus l'erreur $\epsilon$ que la valeur vraie $x_0$.

4. Il convient aussi de les distinguer du *biais*, autre notion statistique dont on parle plus bas.

5. Les expériences de psychologie expérimentale indiquent une incertitude de ce processus de l'ordre de 50 ms [6].

6. Cette hypothèse est à discuter : peut-on supposer les frottements nuls et donc utiliser des périodes successives ? Doit-on faire la mesure sur la première oscillation en s'assurant que les conditions initiales sont identiques ?

## 1.2 Erreurs systématiques

Supposons maintenant qu'on mesure la période $T$ avec un chronomètre faussé qui indique toujours des temps 2% trop faibles. L'étude statistique ne le détectera pas. On parle ici d'*erreur systématique* ou encore de *biais* : c'est la composante de l'erreur qui ne varie pas lorsqu'on répète une mesure dans des conditions identiques. Dans la table 2, on donne un exemple de série de mesures de la période $T$ du pendule, entachées d'erreurs aléatoires, mais aussi d'une erreur systématique additive [7] de $+0,3\,\mathrm{s}$, alors que la valeur vraie est $t_0 = 2\,\mathrm{s}$. On constate que les mesures se répartissent bien aléatoirement autour d'une valeur différente de $t_0$.

TABLE 2 – Série de mesures d'une période d'oscillation d'un pendule de période $t_0 = 2\,\mathrm{s}$, effectuées avec un chronomètre manuel précis au centième de seconde, avec une erreur systématique additive de $+0,3\,\mathrm{s}$. La moyenne des mesures est $\langle t_i \rangle = 2,29\,\mathrm{s}$.

| $i$ | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 |
|---|---|---|---|---|---|---|---|---|---|---|
| $t_i$ (s) | 2,26 | 2,27 | 2,21 | 2,32 | 2,29 | 2,41 | 2,16 | 2,26 | 2,37 | 2,33 |

Les erreurs systématiques ont en réalité des origines diverses :

▶ Erreur d'étalonnage

*Exemple :* Millikan a trouvé une valeur inexacte de la charge $e$ de l'électron parce qu'il avait pris une valeur erronée de la viscosité de l'air. [8]

▶ Oubli d'un paramètre

*Exemple :* Influence de la température sur la vitesse du son. Remarquons au passage que, dans cette mesure, si l'on ne précise pas la température il est impossible de comparer la mesure à une valeur de référence, qui, elle, est donnée pour une certaine température, voire pour un ensemble de températures.

▶ Procédure erronée

*Exemple :* Mesure d'une résistance $R$ à l'aide d'un ampèremètre et d'un voltmètre sans tenir compte des résistances de ces appareils. On rappelle par exemple que dans le montage courte dérivation (figure 1, gauche), la résistance mesurée $R_m = RR_V/(R+R_V)$ dépend de la résistance $R_V$ du voltmètre, et que dans le montage longue dérivation (figure 1, droite) elle dépend de la résistance $R_A$ de l'ampèremètre selon $R_m = R + R_A$ [9].

Les erreurs systématiques sont difficiles à détecter a priori, mais une fois détectées, on peut souvent les corriger (par exemple en tenant compte des résistances de l'ampèremètre et du voltmètre lors de la mesure d'une résistance).

## 1.3 Rôles respectifs des erreurs aléatoires et systématiques

On représente classiquement les rôles respectifs des erreurs aléatoires et systématiques par une analogie avec un tir sur cible (cf. figure 2), le centre de la cible représentant la valeur vraie de la grandeur à mesurer :

7. Dans le cas d'un problème d'étalonnage, pour lequel les temps indiqués sont par exemple toujours 2% trop faibles, on parlera de biais systématique multiplicatif.

8. Il est d'ailleurs intéressant de constater que cette erreur a perduré assez longtemps, ce qui souligne les biais de confirmation qui peuvent affecter la recherche (voir `https://en.wikipedia.org/wiki/Oil_drop_experiment`, section `Millikan's experiment as an example of psychological effects in scientific methodology`.

9. On néglige dans les deux cas les résistances de contact des appareils de mesure. Pour la mesure de faibles résistances, il faudra avoir recours à un montage dit "4 points" ou "4 fils".

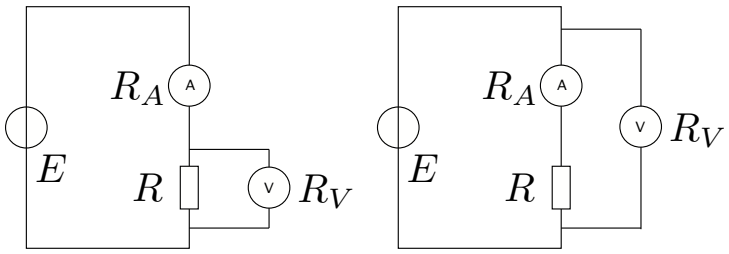


FIGURE 1 – Montages courte dérivation (à gauche) et longue dérivation (à droite)

▶ Tous les impacts sont proches du centre : faibles erreurs aléatoires et faible erreur systématique. Il s'agit du cas (a) de la figure 2. La mesure est à la fois *juste* et *fidèle*[10].

▶ Les impacts sont très étalés mais centrés en moyenne sur le centre de la cible : fortes erreurs aléatoires et faible erreur systématique. Il s'agit du cas (b) de la figure 2. La mesure est *juste* mais peu *fidèle*.

▶ Les impacts sont groupés mais loin du centre : faibles erreurs aléatoires et forte erreur systématique. Il s'agit du cas (c) de la figure 2. La mesure est *fidèle* mais peu *juste*.

▶ Les impacts sont étalés et loin du centre : fortes erreurs aléatoires et forte erreur systématique. Il s'agit du cas (d) de la figure 2. La mesure n'est ni *juste*, ni *fidèle*.

Cet exemple montre par ailleurs que les erreurs systématiques et aléatoires sont généralement présentes simultanément. Anticipant sur les notations de la section 2, en termes probabilistes, on a $\mathsf{E}[X] = x_0$ dans le cas d'erreurs purement aléatoires, et $\mathsf{E}[X] \neq x_0$ si une erreur systématique est présente. L'espérance mathématique $\mathsf{E}$ est l'opérateur moyenne. On appelle *biais* l'écart $b = \mathsf{E}[X] - x_0$ entre la moyenne des mesures et la valeur vraie, et si l'on note $\sigma$ une évaluation de la dispersion des mesures autour de leur moyenne, on distinguera les cas (a) et (b), pour lesquels $b \ll \sigma$, des cas (c) et (d), pour lesquels $b \gtrsim \sigma$.

Le défaut de cette analogie est qu'en général, dans les mesures physiques, on ne connaît bien évidemment pas le biais[11] ! Il doit être évalué en analysant au mieux les sources d'erreurs systématiques. Pour prendre un exemple plus proche de l'expérience en TP, si l'on mesure une distance avec une règle en métal souple tenue à bout de bras, la flexion de la règle va introduire une erreur systématique, car la distance lue est toujours trop grande par rapport à la distance vraie, et une erreur aléatoire car la flexion de la règle est variable d'une mesure à l'autre.

## 1.4 Incertitude

L'incertitude $\delta x$ sur la mesure $x$ de la grandeur aléatoire $X$ traduit les tentatives scientifiques pour estimer l'importance de l'erreur aléatoire. En supposant qu'il n'y a pas de source d'erreur systématique, elle définit un intervalle $[x - \delta x, x + \delta x]$, autour de la valeur mesurée $x$, qui inclut la valeur vraie $x_0$ avec une probabilité donnée. En termes probabilistes, on écrit cette assertion sous la forme $\mathsf{P}(|x - x_0| \leqslant \delta x) = \mathsf{P}_0$, où $0 < \mathsf{P}_0 < 1$ est un nombre donné à l'avance. Notons bien que si une erreur systématique biaise la mesure, il est possible que la valeur vraie ne soit

10. On verra parfois le terme *fiable* à la place de *juste* et *précise* à la place de *fidèle*. La terminologie des programmes est celle donnée ici.

11. Dans notre analogie de tir sur cible, on ne connaît pas la position de la cible...

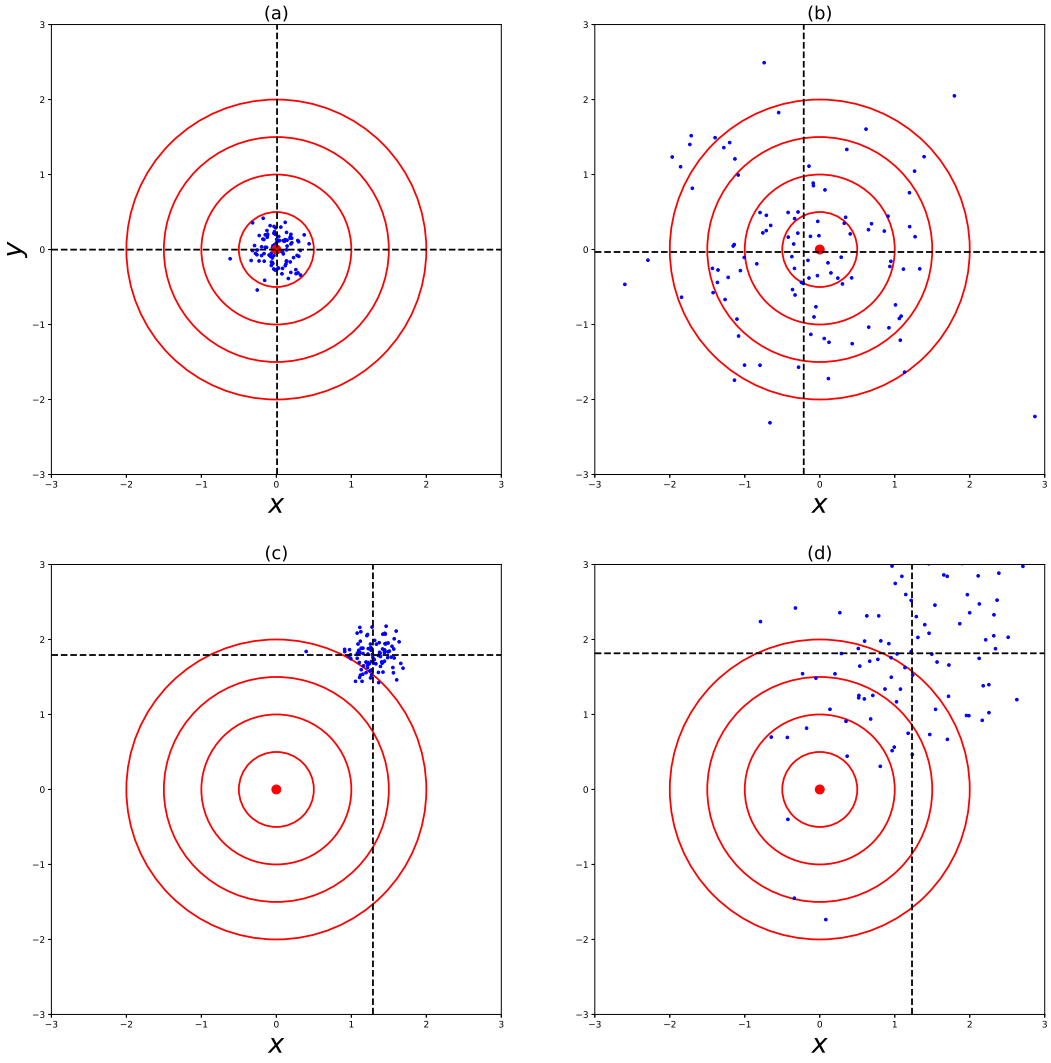


FIGURE 2 – Rôles respectifs des erreurs aléatoires et systématiques dans le cas d'une mesure bidimensionnelle $(x, y)$. La valeur vraie est $(x_0, y_0) = (0, 0)$ au centre des axes marqués en tirets. Les cercles permettent de constater l'erreur des mesures (points bleus) par rapport à cette valeur vraie. Les erreurs aléatoires sont faibles dans les cas (a) et (c), fortes dans les cas (b) et (d). Les erreurs systématiques sont faibles dans les cas (a) et (b), fortes dans les cas (c) et (d). Sur chaque panneau, la position moyenne des points bleus est à l'intersection des lignes en tirets.

pas contenue dans cet intervalle, même si l'on prend $\mathsf{P}_0$ proche de l'unité [12]. L'incertitude n'a trait qu'à la partie aléatoire de l'erreur. Notons qu'elle dépend de la valeur choisie de $\mathsf{P}_0$, qu'on appellera *niveau de confiance*. L'intervalle correspondant $[x - \delta x, x + \delta x]$ est appelé *intervalle de confiance*.

La détermination de l'incertitude sur une mesure n'est a priori pas simple car il faudrait en principe connaître parfaitement la distribution de probabilité de $X$ (voir la section 2), mais on rencontre en pratique deux situations, selon qu'on peut ou non répéter la mesure dans des conditions identiques. Dans le premier cas, c'est-à-dire s'il est possible de recourir à plusieurs mesures de $X$ dans des conditions identiques, on peut procéder à une évaluation statistique de l'incertitude [13]. On parle alors d'évaluation de type A. Cette approche, qu'on appelle *approche statistique*, fait l'objet de la section 3.

Si l'on ne dispose pas du temps nécessaire pour faire une série de mesures, on procède à une évaluation de type B, qui consiste à estimer $\delta x$ à partir des spécifications des appareils de mesure et des conditions expérimentales. Cela revient à supposer une loi de probabilité des mesures, c'est pourquoi on appelle cette approche, discutée à la section 4.6, *approche probabiliste*.

# 2 Rappels élémentaires de probabilités et statistiques

L'incertitude étant une traduction mathématique du caractère aléatoire de la répartition des mesures d'une grandeur physique $X$, il est important d'adopter le langage des probabilités et des statistiques pour aborder correctement les calculs d'incertitudes qui suivent. On rappelle donc dans cette section 2 quelques éléments utiles. En première lecture, et si on n'est pas trop à l'aise avec les calculs, on peut sauter cette section.

## 2.1 Variables aléatoires

En toute rigueur, une variable aléatoire $X$ est une application d'un espace probabilisé $\mathcal{C}$ (essentiellement défini comme un ensemble d'*évènements* possibles et muni d'une mesure $\mathsf{P}$ appelée *probabilité*), vers l'ensemble des réels $\mathbb{R}$. Ceci permet d'assigner un réel $x$ au résultat $\omega$ d'une expérience, lequel résultat peut être non-numérique comme "pile" ou "face" par exemple. On écrit alors $X(\omega) = x$. Notons que dans le cas de grandeurs physiques, l'ensemble des résultats possibles est déjà numérique, ce qui rend cette subtilité superflue. Notons aussi que les variables aléatoires peuvent prendre des valeurs continues $x \in \mathbb{R}$, ou des valeurs discrètes $\{x_i\}_{i\in\mathbb{N}}$ (lors d'un lancer de dé par exemple, ou lorsqu'on cherche le nombre de monomères dans une chaîne polymère, ou le nombre de photons dans une cavité résonnante...).

## 2.2 Fonction de répartition

On définit la *fonction de répartition* [14] $F$ de la variable aléatoire $X$ comme [15]

$$F(x) = \mathsf{P}\left(X \leqslant x\right).$$

En termes de mesure d'une grandeur physique $X$, elle donne, pour chaque réel $x$, la probabilité que le résultat d'une mesure de $X$ soit inférieur ou égal à $x$. Bien entendu, cette fonction est

12. Elle le sera néanmoins évidemment si $\mathsf{P}_0 = 1$, mais dans ce cas, l'incertitude croît jusqu'à ce que l'intervalle $[x - \delta x, x + \delta x]$ couvre tout le domaine de valeurs que peut prendre la mesure, ce qui n'a pas d'intérêt.

13. Cela revient à construire un histogramme des mesures obtenues, ce qui correspond à une approximation finie de la distribution de probabilité de $X$.

14. On l'appelle parfois aussi fonction de distribution cumulative (CDF pour *Cumulative Distribution Function*).

15. Plus précisément, il s'agit de la probabilité - en tant que mesure sur $\mathcal{C}$ - de l'ensemble $\Omega$ des évènements $\omega$ tels que $X(\omega) \leqslant x$.

monotone croissante, et on a les limites suivantes (en faisant l'hypothèse que les valeurs possibles de la mesure de $X$ couvrent l'ensemble des réels, de $-\infty$ à $+\infty$) :

$$\lim_{x\to-\infty} F(x) = 0 \qquad \text{et} \qquad \lim_{x\to+\infty} F(x) = 1.$$

La figure 3 donne un exemple de fonction de répartition d'une variable aléatoire continue (à gauche) et un exemple de fonction de répartition d'une variable aléatoire discrète (à droite), tirés de [2].

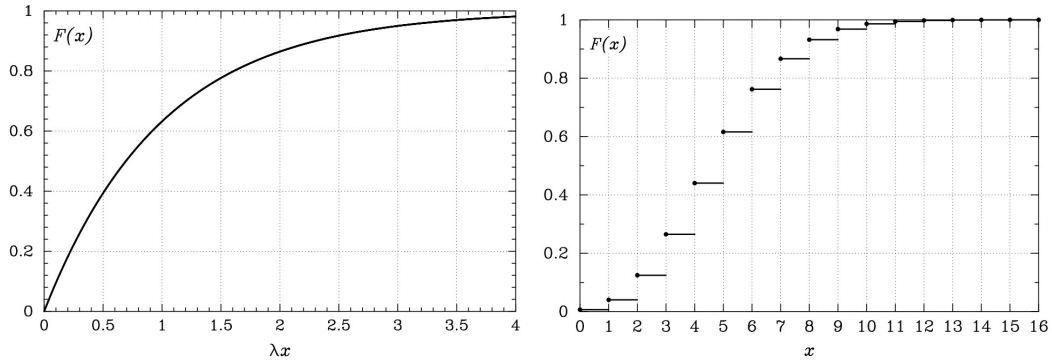


FIGURE 3 – Exemples, tirés de [2], de fonctions de répartition d'une variable aléatoire continue (à gauche) et discrète (à droite).

Notons que la fonction de répartition peut être discontinue, si la variable aléatoire peut prendre des valeurs discrètes. Comme on peut montrer que $F(x^+) - F(x^-) = \mathsf{P}(X = x)$, on en déduit que pour une variable aléatoire à valeurs continues, la probabilité que $X$ prenne exactement la valeur $x$ est nulle. On se placera dans ce cas continu pour la suite, sauf mention contraire.

## 2.3 Densité de probabilité

On appelle *densité de probabilité* ou *distribution de probabilité*[16] d'une variable aléatoire $X$ à valeurs continues la fonction $f(x)$ telle que la probabilité pour que $X$ soit comprise entre $x$ et $x + \mathrm{d}x$ s'écrive $\mathsf{P}(x < X \leqslant x + \mathrm{d}x) = f(x)\mathrm{d}x$. Autrement dit, $f$ est la dérivée de la fonction de répartition $F$, soit $f(x) = F'(x)$. Il va de soi que $f$ est donc positive et normalisée :

$$f(x) = F'(x) \geqslant 0 \qquad \text{et} \qquad \int_{-\infty}^{\infty} f(x)\mathrm{d}x = F(+\infty) - F(-\infty) = 1$$

De même, la probabilité pour que la variable aléatoire $X$ prenne une valeur comprise entre $a$ et $b$ est bien entendu donnée par

$$\mathsf{P}(a < X \leqslant b) = \int_a^b f(x)\mathrm{d}x = F(b) - F(a)$$

Dans le cas de mesures de grandeurs physiques, on retiendra que $f(x)\mathrm{d}x$ est la probabilité pour que la mesure obtenue soit comprise entre $x$ et $x + \mathrm{d}x$. Les figures 4 et 5 donnent des exemples de distributions de probabilité de variables continues, les figures 6 et 7 donnent des exemples de distributions de probabilité de variables discrètes.

16. PDF pour *Probability Distribution Function*.

## 2.4 Cas de variables aléatoires à plusieurs dimensions

Pour des expériences dont le résultat est $n$-dimensionnel, on généralise les définitions précédentes en posant le vecteur aléatoire $\mathbf{X} = (X_1, \ldots, X_n)$ où chacun des $X_i$ est une variable aléatoire réelle. On définit ainsi la fonction de répartition

$$F(x_1, \ldots, x_n) = \mathsf{P}(X_1 \leqslant x_1, \ldots, X_n \leqslant x_n)$$

qui est la probabilité pour que la mesure de chacune des variables $X_i$ donne un résultat inférieur ou égal à $x_i$. La fonction de répartition est à valeurs dans $[0, 1]$, croissante en chacune des variables, égale à 0 si l'une quelconque de celles-ci vaut $-\infty$ et égale à 1 si elles valent toutes $+\infty$. On définit ensuite - supposant qu'elle existe - la densité de probabilité $n$-dimensionnelle

$$f(x_1, \ldots, x_n) = \frac{\partial^n F}{\partial x_1 \ldots \partial x_n}$$

qui est positive et d'intégrale unité sur $\mathbb{R}^n$. Notons que si les variables aléatoires $X_i$ sont indépendantes les unes des autres, on peut factoriser la fonction de répartition et la distribution de probabilité $n$-dimensionnelles

$$F(x_1, \ldots, x_n) = F_1(x_1) \ldots F_n(x_n) \qquad \text{et} \qquad f(x_1, \ldots, x_n) = f_1(x_1) \ldots f_n(x_n).$$

## 2.5 Changement de variable aléatoire

Il est fréquent que le résultat d'une mesure ne donne pas directement la grandeur cherchée. Par exemple, on cherche à mesurer une résistance $R = U/I$ à partir d'une mesure d'intensité $I$ à l'ampèremètre et d'une mesure de tension $U$ au voltmètre.

Avant de traiter ce cas, il faut d'abord comprendre le cas d'une variable aléatoire $Y$ déduite d'une autre $X$ via une fonction $h$ de cette seule variable, soit $Y = h(X)$. Par exemple, on pourra chercher à estimer l'incertitude sur le carré de l'intensité $I^2$, sachant qu'on mesure l'intensité $I$. La question qui se pose, étant donné que les erreurs sur la mesure de $X$ mènent à des erreurs sur la grandeur cherchée $Y$, est de savoir comment se propagent les incertitudes. En termes mathématiques, il s'agit de calculer la densité de probabilité $g$ de $Y$ connaissant celle $f$ de $X$. On montre que

$$g(y) = \sum_k \frac{f(x_k)}{|h'(x_k)|}$$

où la somme est étendue à tous les réels $x_k$ solutions de $h(x_k) = y$. Par exemple, si $Y = X^2$, on a les deux solutions $x_+ = \sqrt{y}$ et $x_- = -\sqrt{y}$, à condition que $y > 0$. On a alors, pour $y > 0$, la distribution de probabilité suivante :

$$g(y) = \frac{f(\sqrt{y})}{|2\sqrt{y}|} + \frac{f(-\sqrt{y})}{|-2\sqrt{y}|} = \frac{f(\sqrt{y}) + f(-\sqrt{y})}{2\sqrt{y}}.$$

Pour $y \leqslant 0$, la distribution de probabilité est évidemment nulle, $g(y) = 0$.

Dans le cas d'un changement de variable $n$-dimensionnel

$$\mathbf{Y} = (Y_1, \ldots, Y_n) = \mathbf{h}(X_1, \ldots, X_n)$$

on a alors la densité de probabilité $g$ de $\mathbf{Y}$ en fonction de celle $f$ de $\mathbf{X}$ par la relation

$$g(y_1, \ldots, y_n) = \sum_k f\left(x_1^{(k)}, \ldots, x_n^{(k)}\right) \left| \frac{\partial(y_1, \ldots, y_n)}{\partial(x_1, \ldots, x_n)} \right|^{-1}_{x_i = x_i^{(k)}}$$

qui fait intervenir le *Jacobien* de la transformation **h**, qui est le déterminant de la matrice constituée par les dérivées partielles,

$$J = \left| \frac{\partial (y_1, \ldots, y_n)}{\partial (x_1, \ldots, x_n)} \right| = \begin{vmatrix} \frac{\partial y_1}{\partial x_1} & \cdots & \frac{\partial y_1}{\partial x_n} \\ \cdots & \cdots & \cdots \\ \frac{\partial y_n}{\partial x_1} & \cdots & \frac{\partial y_n}{\partial x_n} \end{vmatrix}$$

On peut alors utiliser ce résultat pour trouver la distribution de probabilité de grandeurs qui sont fonction de plusieurs autres grandeurs. Par exemple, pour trouver la distribution de probabilité de la somme de deux variables aléatoires $X_1$ et $X_2$ supposées indépendantes, on pose le changement de variable

$$(X_1, X_2) \mapsto (Y_1 = X_1, Y_2 = X_1 + X_2).$$

On a alors le Jacobien

$$J = \left| \frac{\partial (y_1, y_2)}{\partial (x_1, x_2)} \right| = \begin{vmatrix} \frac{\partial y_1}{\partial x_1} & \frac{\partial y_1}{\partial x_2} \\ \frac{\partial y_2}{\partial x_1} & \frac{\partial y_2}{\partial x_2} \end{vmatrix} = \begin{vmatrix} 1 & 0 \\ 1 & 1 \end{vmatrix} = 1$$

et donc la distribution de probabilité jointe de $(Y_1, Y_2)$ comme $g(y_1, y_2) = f(y_1, y_2 - y_1)$. Pour obtenir la distribution de probabilité de la seule somme $Y_2 = X_1 + X_2$, on *marginalise* sur la variable dont le résultat ne nous intéresse pas, à savoir $Y_1$. Ce terme de marginalisation revient à dire qu'on intègre sur $y_1$, soit

$$g(y_2) = \int_{-\infty}^{+\infty} f(y_1, y_2 - y_1) \mathrm{d}y_1.$$

Dans le cas de variables aléatoires indépendantes, la densité de probabilité $f$ est factorisable, et on a donc

$$g(y_2) = \int_{-\infty}^{+\infty} f_{X_1}(y_1) f_{X_2}(y_2 - y_1) \mathrm{d}y_1$$

On constate ainsi que $g$ est le produit de convolution des distributions de probabilités $f_{X_1}$ et $f_{X_2}$. On procède de manière semblable pour obtenir la distribution de probabilité de la différence, du produit et du quotient de $X_1$ et $X_2$. Finalement, en supposant toujours que les variables aléatoires $X_1$ et $X_2$ sont indépendantes, on a :

- Pour $Y = X_1 + X_2 : g(y) = \displaystyle\int_{-\infty}^{+\infty} f_{X_1}(u) f_{X_2}(y - u) \mathrm{d}u$
- Pour $Y = X_1 - X_2 : g(y) = \displaystyle\int_{-\infty}^{+\infty} f_{X_1}(u) f_{X_2}(y + u) \mathrm{d}u$
- Pour $Y = X_1 X_2 : g(y) = \displaystyle\int_{-\infty}^{+\infty} f_{X_1}(u) f_{X_2}\left(\frac{y}{u}\right) \frac{1}{|u|} \mathrm{d}u$
- Pour $Y = X_2 / X_1 : g(y) = \displaystyle\int_{-\infty}^{+\infty} f_{X_1}(u) f_{X_2}(uy) |u| \mathrm{d}u$

## 2.6 Espérance d'une variable aléatoire

La *valeur moyenne* ou *espérance* $\mathsf{E}[X]$ d'une grandeur aléatoire $X$ à valeurs continues possédant une densité de probabilité $f$ est donnée par l'intégrale

$$\mathsf{E}[X] = \int_{-\infty}^{\infty} x f(x) \mathrm{d}x.$$

On peut montrer que l'opérateur valeur moyenne est linéaire

$$\mathsf{E}[a_1 X_1 + \ldots + a_n X_n] = a_1 \mathsf{E}[X_1] + \ldots + a_n \mathsf{E}[X_n]$$

et que pour deux variables aléatoires indépendantes $X$ et $Y$, on a $\mathsf{E}[XY] = \mathsf{E}[X]\,\mathsf{E}[Y]$.

Quand on considère un changement de variable aléatoire $Y = h(X)$, on a

$$\mathsf{E}[Y] = \int_{-\infty}^{\infty} h(x) f(x) \mathrm{d}x$$

ce qui implique qu'il n'est pas nécessaire de connaître la distribution de probabilité de $Y$ pour en calculer la moyenne, dès lors qu'on connaît la densité de probabilité de $X$. Ceci sera important pour calculer les incertitudes sur une grandeur déduite d'une autre directement mesurée.

## 2.7 Moments

### 2.7.1 Définition

Les moments sont des nombres permettant de caractériser la distribution de probabilité d'une variable aléatoire $X$, à défaut de connaître complètement sa densité de probabilité $f$. En fait, la connaissance de l'ensemble des moments est équivalente à celle de $f$.

On définit le moment non-centré d'ordre $k \in \mathbb{N}$ par

$$\mu'_k = \mathsf{E}\left[X^k\right] = \int_{-\infty}^{\infty} x^k f(x) \mathrm{d}x$$

ainsi $\mu'_1$ est la moyenne de $X$. On définit ensuite le moment centré d'ordre $k \in \mathbb{N}$ par

$$\mu_k = \mathsf{E}\left[(X - \mathsf{E}[X])^k\right] = \int_{-\infty}^{\infty} (x - \mu'_1)^k f(x) \mathrm{d}x$$

On peut montrer que les moments et la densité de probabilité sont reliés par

$$Z(u) = \int_{\mathbb{R}} e^{iux} f(x) \mathrm{d}x = \sum_{k \geqslant 0} \mu'_k \frac{(iu)^k}{k!}$$

la fonction $Z(u)$ étant appelée *fonction caractéristique*. On voit que les moments sont donc directement liés aux coefficients du développement en série entière de la transformée de Fourier de la distribution de probabilité.

### 2.7.2 Variance et écart-type

Parmi les moments, il convient de souligner l'importance de certains d'entre eux. On a déjà dit que la moyenne de $X$ était égale au moment non-centré d'ordre un, $\mathsf{E}[X] = \mu'_1$. On appelle *variance* le moment centré d'ordre deux, qui est positif et qu'on note souvent $\sigma^2$, soit

$$\sigma^2 = \mu_2 = \int_{-\infty}^{\infty} (x - \mathsf{E}[X])^2 f(x) \mathrm{d}x$$

C'est une mesure de la dispersion des valeurs de $X$ autour de sa valeur moyenne. Plus les valeurs de $X$ se concentrent autour de la moyenne, plus la variance est faible. On définit ainsi l'écart-type $\sigma$, racine carrée de la variance, qui est positif par convention, et qui donne *grosso modo* l'étendue de la plage de valeurs qu'on peut attendre pour les mesures de $X$. Il peut donc jouer le rôle d'une incertitude $\delta x$. On parlera ainsi d'*incertitude-type* si elle correspond à un écart-type, soit $\delta x = \sigma$, mais on rencontre aussi l'*incertitude élargie*[17], pour laquelle $\delta x = 2\sigma$.

Note : quand la grandeur aléatoire $X$ ne prend que des valeurs discrètes $x_i$ avec des probabilités respectives $p_i$, les moments de la distribution sont donnés par

$$\mu'_k = \sum_i x_i^k p_i \qquad \text{et} \qquad \mu_k = \sum_i \left(x_i - \mu'_1\right)^k p_i$$

En particulier, la valeur moyenne et la variance de la distribution sont respectivement :

$$\mathsf{E}\left[X\right] = \sum_i x_i p_i \qquad \text{et} \qquad \sigma^2 = \sum_i \left(x_i - \mathsf{E}\left[X\right]\right)^2 p_i$$

### 2.7.3 Moments d'ordre plus élevé

Pour préciser plus avant la distribution de $X$, on peut calculer les moments centrés d'ordre trois et quatre, et en déduire le coefficient d'asymétrie (*skewness*) $\gamma_1$ et le coefficient d'aplatissement (*kurtosis*) $\gamma_2$ via

$$\gamma_1 = \frac{\mu_3}{\mu_2^{3/2}} \qquad \text{et} \qquad \gamma_2 = \frac{\mu_4}{\mu_2^2} - 3$$

Le coefficient $\gamma_1$ est nul pour une distribution symétrique par rapport à $\mu'_1$. D'autre part, ainsi défini, le coefficient $\gamma_2$ est nul pour la distribution Gaussienne (voir 2.8.3).

## 2.8 Exemples de distributions de probabilité

### 2.8.1 Distribution uniforme (ou rectangulaire)

On considère une variable aléatoire $X$ de densité de probabilité $\Pi(x)$ uniforme non nulle entre $\mu - \Delta$ et $\mu + \Delta$, et nulle partout ailleurs. Par normalisation, cette densité de probabilité $\Pi(x)$ vaut $1/(2\Delta)$ sur cet intervalle $[\mu - \Delta, \mu + \Delta]$ (voir la figure 4).

La moyenne est alors égale à $\mu$ et les moments centrés de la distribution s'écrivent :

$$\mu_k = \int_{\mu-\Delta}^{\mu+\Delta} \frac{(x-\mu)^k}{2\Delta} \mathrm{d}x = \int_{-\Delta}^{+\Delta} \frac{x^k}{2\Delta} \mathrm{d}x = \frac{1}{2\Delta} \frac{\Delta^{k+1} - (-\Delta)^{k+1}}{k+1}$$

On en déduit que

$$\mu_{2p} = \frac{\Delta^{2p}}{2p+1} \qquad \text{et} \qquad \mu_{2p+1} = 0$$

Cela sera utile au 4.6.

17. Cette notion d'incertitude élargie est parfois utilisée pour des intervalles plus larges, notamment $\delta x = 3\sigma$. Il convient donc de faire attention à ce qui est indiqué sur les notices.

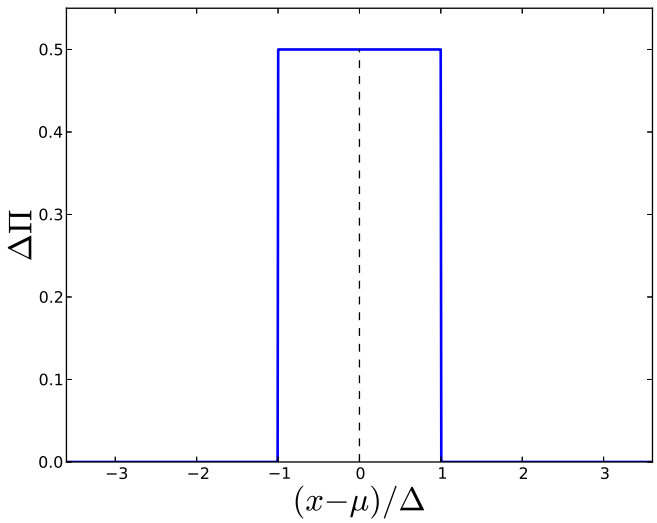


FIGURE 4 – Distribution uniforme normalisée $\Delta \times \Pi(x)$, représentée en fonction de la variable normalisée $(x-\mu)/\Delta$. La valeur moyenne nulle est représentée par les tirets noirs.

### 2.8.2 Distribution triangulaire

Une variable aléatoire $X$ suit une loi triangulaire sur l'intervalle $[a, b]$ si sa fonction de distribution en probabilité est nulle hors de cet intervalle et a la forme suivante sinon :

$$f(x) = \frac{4(x-a)}{(b-a)^2} \quad \text{pour } a \leqslant x \leqslant \frac{a+b}{2} \qquad f(x) = \frac{4(b-x)}{(b-a)^2} \quad \text{pour } \frac{a+b}{2} \leqslant x \leqslant b$$

La moyenne et l'écart-type de cette distribution sont respectivement

$$\mathsf{E}[X] = \frac{a+b}{2} \qquad \sigma = \frac{b-a}{2\sqrt{6}}$$

On pourra supposer une telle distribution au 4.6.

TABLE 3 – Probabilité $\mathsf{P}$ qu'une mesure de $X$ se trouve dans l'intervalle $[\mu - \alpha\sigma, \mu + \alpha\sigma]$ si $X$ suit une loi normale de moyenne $\mu$ et d'écart-type $\sigma$.

| $\alpha$ | 0,5 | 1 | 1,5 | 2 | 2,5 | 3 | 3,5 | 4 |
|---|---|---|---|---|---|---|---|---|
| $\mathsf{P}$ (%) | 38 | 68,3 | 87 | 95,4 | 98,8 | 99,7 | 99,95 | 99,99 |

### 2.8.3 Distribution de Gauss ou loi normale

La densité de probabilité Gaussienne de valeur moyenne $\mu$ et d'écart-type $\sigma$ s'écrit :

$$\mathcal{N}_{\mu,\sigma}(x) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left[-\frac{(x-\mu)^2}{2\sigma^2}\right].$$

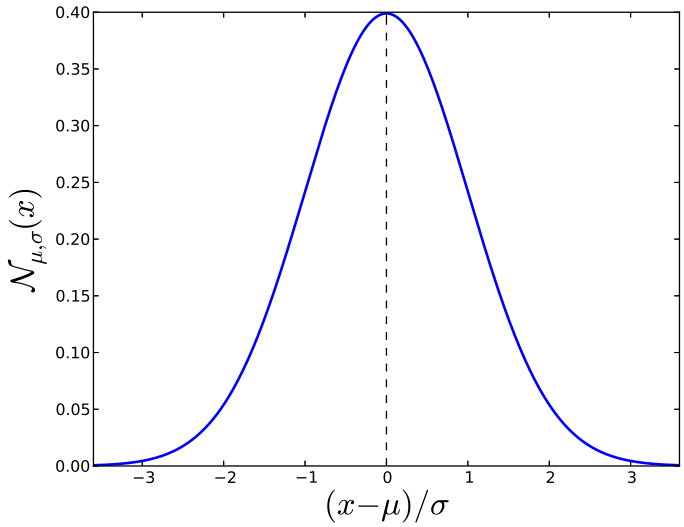


FIGURE 5 – Distribution Gaussienne $\mathcal{N}_{\mu,\sigma}$ représentée en fonction de la variable normalisée $(x-\mu)/\sigma$. La valeur moyenne nulle est représentée par les tirets noirs. L'aire sous la courbe comprise entre $-1$ et $+1$ vaut 0,683 et celle entre $-2$ et $+2$ vaut 0,954.

C'est une distribution symétrique autour de la moyenne dont le graphe est présenté sur la figure 5 en fonction de $z = (x-\mu)/\sigma$. L'intégrale de $\mathcal{N}_{\mu,\sigma}$ entre $\mu-\sigma$ et $\mu+\sigma$ vaut 0,683, c'est à dire que la probabilité que la mesure $x$ de $X$ soit à moins d'un écart-type de la moyenne (en valeur absolue) est de 68,3 %. Autrement dit, le niveau de confiance $\mathsf{P}_0$ associé à l'incertitude-type pour une distribution Gaussienne est $\mathsf{P}_0 = 0,683$. Cette probabilité passe à 95.4% pour l'incertitude-élargie (2 écarts-types) et à 99,7 % pour un écart à la moyenne de 3 écarts-types (voir le tableau 3).

L'importance physique de la distribution Gaussienne est discutée au paragraphe 2.9.

### 2.8.4 Distribution binômiale

On réalise $n$ fois une expérience n'ayant que deux résultats possibles, l'un de probabilité $p$, l'autre de probabilité $1-p$. La probabilité d'obtenir $k$ fois le résultat de probabilité $p$ est donnée par la distribution binômiale :

$$B_{n,p}(k) = C_n^k p^k (1-p)^{n-k} \qquad \text{avec} \qquad C_n^k = \frac{n!}{k!(n-k)!}.$$

Cette distribution de probabilité, représentée sur la figure 6, et définie uniquement pour les valeurs de $k$ entières, a pour moyenne $np$ et pour écart-type $\sqrt{np(1-p)}$. Elle n'est pas symétrique autour de sa moyenne (car définie uniquement pour $k \geqslant 0$). Néanmoins, si $n$ est grand et si les probabilités $p$ et $1-p$ ne sont pas trop voisines de 0, la loi normale de moyenne $\mu = np$ et d'écart-type $\sigma = \sqrt{np(1-p)}$ est une très bonne approximation de la distribution binômiale. Dans la pratique, cette approximation est très convenable dès que $np$ et $n(1-p)$ sont tous deux supérieurs à 5. Il faut cependant prendre garde que la comparaison entre les deux n'est pas directe, car l'une porte sur une variable aléatoire discrète et l'autre sur une variable aléatoire continue.

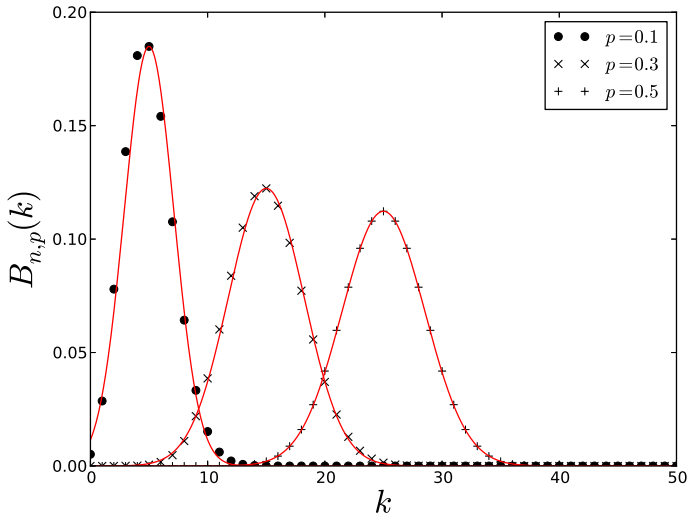


FIGURE 6 – Distribution binômiale $B_{n,p}(k)$ en fonction de $k$, pour différentes valeurs de $p$ (indiquées en légende), et $n = 50$. Les courbes montrent les distributions Gaussiennes approchant les distributions binômiales correspondantes.

### 2.8.5 Distribution de Poisson

On la rencontre dans des situations où des évènements rares et indépendants se produisent aléatoirement dans le temps avec une probabilité par unité de temps connue, par exemple lorsqu'on compte, pendant un temps donné, un nombre de désintégrations radioactives, ou le nombre de photons $\gamma$ reçus sur un détecteur, ou encore lorsqu'on décrit le nombre de molécules se trouvant à l'intérieur d'un petit volume de gaz. Si les évènements sont indépendants les uns des autres, le nombre $n$ d'évènements qui se produisent pendant une durée fixe est une grandeur aléatoire dont la distribution de probabilité est la loi de Poisson :

$$P_N(n) = \frac{N^n}{n!} e^{-N}$$

$N$ est le nombre moyen d'évènements observés pendant la durée donnée :

$$\mathsf{E}[n] = \sum_{n \geqslant 0} n P_N(n) = N e^{-N} \sum_{n \geqslant 1} \frac{N^{n-1}}{(n-1)!} = N e^{-N} e^{N} = N.$$

La distribution de Poisson n'est pas symétrique autour de sa moyenne (voir la figure 7). Une de ses particularités est qu'elle ne dépend que d'un paramètre. Son écart-type est directement relié à sa moyenne par $\sigma = \sqrt{N}$.

On l'écrit souvent en notant $\lambda$ la probabilité d'un évènement par unité de temps, de sorte que $N = \lambda T$ avec $T$ la durée d'observation. On écrit alors la distribution de Poisson sous la forme

$$P_\lambda(n) = \frac{(\lambda T)^n}{n!} e^{-\lambda T}.$$

La distribution de Poisson de valeur moyenne $N = np$ est une très bonne approximation de la distribution binômiale $B_{n,p}$, lorsqu'on compte beaucoup d'évènements rares, c'est à dire pour les

grandes valeurs de $n$ et les faibles valeurs de $p$. On pourra comparer par exemple la courbe pour $p = 0.1$ sur la Fig. 6 avec le cas $N = 5$ sur la Fig. 7. Comme la loi binômiale, la distribution de Poisson est très proche d'une Gaussienne dès que le nombre d'évènements observés est grand.

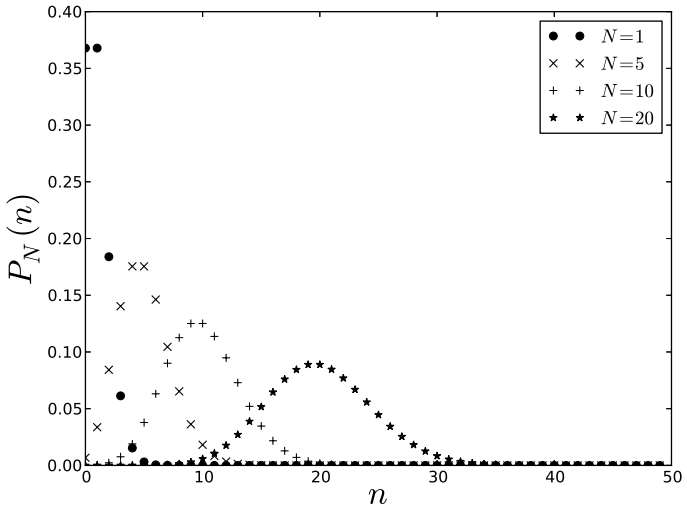


FIGURE 7 – Distribution de Poisson $P_N(n)$ en fonction de $n$, pour différentes valeurs de $N$.

### 2.9 Importance physique de la loi normale

Dans la pratique, si l'on répète un grand nombre de fois une mesure physique, on obtient le plus souvent pour les résultats de la mesure une distribution de probabilité Gaussienne. On a d'ailleurs déjà constaté cette tendance sur les distributions binômiale et Poissonienne (Figures 6 et 7). On peut en fait démontrer ce résultat pour toutes les mesures sujettes à des erreurs systématiques négligeables mais à de nombreuses sources d'erreurs aléatoires indépendantes, et ce quelles que soient les formes des distributions de ces erreurs. La version mathématique de ce théorème est appelée *théorème central sur les limites*. On en trouvera une illustration physique simple dans [1], §10.5, p. 235.

## 3 Evaluation des incertitudes par des méthodes statistiques

On présente ici l'évaluation de type A d'une incertitude expérimentale, c'est-à-dire lorsqu'on dispose de mesures répétées $x_1, \ldots, x_n$ d'une grandeur physique $X$ et qu'une étude statistique est donc possible. On considère d'abord le cas où l'on a accès directement à la grandeur physique recherchée, puis on donne les méthodes de calcul d'une incertitude dans le cas où une formule mathématique relie la grandeur physique d'intérêt aux diverses grandeurs effectivement mesurées.

### 3.1 Analyse statistique d'une série de mesures

On s'occupe ici de la mesure d'une grandeur physique $X$ dont les sources de variabilité sont uniquement aléatoires. On cherche à caractériser la distribution de probabilité $f$ de $X$, en évaluant

le mieux possible la valeur moyenne $\overline{x} = \mathsf{E}[X]$ et l'écart-type $\sigma$ de cette distribution, ce qui suffit si la distribution de probabilité est Gaussienne, mais est une grossière simplification dans le cas contraire. Le traitement statistique est fondé sur la répétition des mesures de $X$. Si l'on était capable de réaliser une infinité de mesures, on déterminerait la distribution de probabilité de $X$, notée $f(x)$, et en particulier sa valeur moyenne $\overline{x} = \mathsf{E}[X]$ (égale à la valeur vraie $x_0$ en l'absence d'erreur systématique, ce qu'on suppose ici) et son écart type $\sigma$. C'est ce qu'illustre la figure 8.

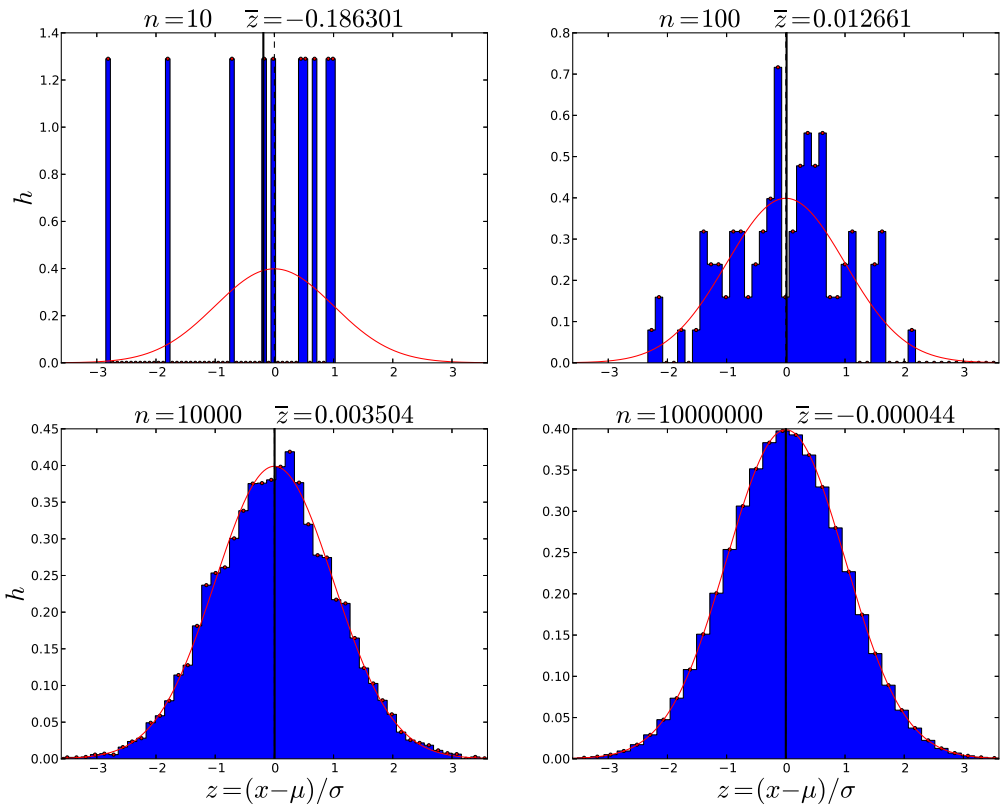


FIGURE 8 – Évolution des histogrammes des mesures d'une variable distribuée selon une loi normale de moyenne nulle et de variance unité, lorsqu'on fait varier le nombre $n$ de points de mesure. Le trait vertical continu représente la moyenne des valeurs mesurées, indiquée au-dessus de chaque figure, et la courbe en trait plein la distribution normale.

Dans la pratique, on réalise un nombre fini $n$ de mesures, de résultats respectifs $x_1$, $x_2$, …, $x_n$, dont on cherche à extraire les meilleures estimations de $x_0$ et $\sigma$. Les méthodes statistiques qui permettent d'obtenir ces meilleures estimations sont présentées dans les paragraphes suivants, avec un minimum de démonstrations.

### 3.1.1 Meilleure estimation de la moyenne

Étant données $n$ mesures $x_1$, $x_2$, …, $x_n$ de $X$, on peut construire la valeur moyenne (appelée *moyenne empirique* ou *moyenne expérimentale*)

$$\overline{x}_n = \frac{x_1 + x_2 + \ldots + x_n}{n} = \frac{1}{n}\sum_{i=1}^{n} x_i$$

qu'on peut voir comme une réalisation de la variable aléatoire

$$\overline{X}_n = \frac{X_1 + X_2 + \ldots + X_n}{n} = \frac{1}{n}\sum_{i=1}^{n} X_i$$

où les $X_i$ sont $n$ variables aléatoires *indépendantes et identiquement distribuées (i.i.d.)* c'est-à-dire qu'elles ont toutes la même densité de probabilité (celle de $X$). Par linéarité, on a $\mathsf{E}\left[\overline{X}_n\right] = \mathsf{E}\left[X\right] = x_0$. On obtient donc ainsi une estimation $\widehat{x_0}$ non biaisée de $x_0$

$$\widehat{x_0} = \overline{x}_n = \frac{1}{n}\sum_{i=1}^{n} x_i.$$

On peut montrer que cette estimation est la meilleure possible. La démonstration de ce résultat est assez simple dans le cas où la grandeur à mesurer a une distribution de probabilité Gaussienne (voir [1], §5.5, p.128 et annexe E1, p.280) : les meilleures estimations pour $x_0$ et $\sigma$ sont en effet celles qui maximisent la probabilité

$$\mathsf{P}_{x_0,\sigma}(x_1, \ldots, x_n) = \frac{1}{(2\pi)^{n/2}\sigma^n} \exp\left[-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(x_i - x_0)^2\right]$$

d'obtenir les résultats $x_1$, $x_2$, …, $x_n$ lors de $n$ mesures. La dérivation de cette probabilité par rapport à $x_0$ (les $x_i$ et $\sigma$ étant fixés) montre que cette dérivée est nulle pour $x_0 = \overline{x}_n$, et donc que la probabilité écrite ci-dessus atteint là son maximum.

### 3.1.2 Meilleure estimation de la variance et de l'écart-type

À partir des mêmes $n$ variables aléatoires i.i.d., on peut construire la *variance empirique*

$$S^2 = \frac{1}{n}\sum_{i=1}^{n}\left(X_i - \overline{X}_n\right)^2.$$

Cependant, l'espérance de cette variable aléatoire n'est pas $\sigma^2$, ce qui implique que $S^2$ est un estimateur *biaisé* de la variance. En effet :

$$\mathsf{E}\left[S^2\right] = \frac{1}{n}\sum_{i=1}^{n}\mathsf{E}\left[\left(X_i - \overline{X}_n\right)^2\right] = \mathsf{E}\left[X^2\right] + \mathsf{E}\left[\overline{X}_n^2\right] - \frac{2}{n}\sum_{i=1}^{n}\mathsf{E}\left[X_i\overline{X}_n\right]$$

étant donné que les variables aléatoires $X_i$ sont identiquement distribuées, ce qui implique que les $X_i\overline{X}_n$ et les $X_i^2$ le sont aussi. Comme elles sont également indépendantes, on a

$$\mathsf{E}\left[\overline{X}_n^2\right] = \mathsf{E}\left[\frac{1}{n^2}\sum_{i=1}^{n}\sum_{j=1}^{n} X_iX_j\right] = \mathsf{E}\left[\frac{1}{n^2}\left\{\sum_{i=1}^{n} X_i^2 + \sum_{i=1}^{n}\sum_{j\neq i} X_iX_j\right\}\right] = \frac{\mathsf{E}\left[X^2\right]}{n} + \frac{n-1}{n}\mathsf{E}\left[X\right]^2$$

et

$$\mathsf{E}\left[X_i\overline{X}_n\right] = \frac{1}{n}\left\{\mathsf{E}\left[X_i^2\right] + \mathsf{E}\left[\sum_{j\neq i} X_iX_j\right]\right\} = \frac{\mathsf{E}\left[X^2\right]}{n} + \frac{n-1}{n}\mathsf{E}\left[X\right]^2$$

on en déduit que

$$\mathsf{E}\left[S^2\right] = \left(1-\frac{1}{n}\right)\left(\mathsf{E}\left[X^2\right] - \mathsf{E}\left[X\right]^2\right) = \left(1-\frac{1}{n}\right)\sigma^2.$$

Ceci montre que la variance empirique est un estimateur biaisé et n'est donc pas la meilleure estimation de la variance vraie. Il vaut mieux considérer la *variance empirique modifiée*

$$S'^2 = \frac{n}{n-1}S^2 = \frac{1}{n-1}\sum_{i=1}^{n}\left(X_i - \overline{X}_n\right)^2 \qquad \text{de sorte que} \qquad \mathsf{E}\left[S'^2\right] = \sigma^2.$$

Ainsi, la meilleure estimation de $\sigma^2$ déduite des $n$ mesures $x_1$, $x_2$, …, $x_n$, notée $\sigma_n^2$, est donnée par la réalisation de la variance empirique modifiée associée à ces mesures :

$$\sigma_n^2 = \frac{1}{n-1}\sum_{i=1}^{n}\left(x_i - \overline{x}_n\right)^2.$$

Le facteur $n-1$, et non pas $n$, vient donc du fait que la formule ci-dessus utilise $\overline{x}_n$, seule quantité accessible à l'expérience, en lieu et place de la valeur vraie $x_0$. Il est aisé de s'en souvenir : on conçoit bien qu'il n'est pas possible d'estimer l'écart-type d'une distribution à partir d'une seule mesure. Notons que si l'on connaissait $x_0$, une estimation non biaisée de la variance de la distribution serait donnée par

$$\sigma_n'^2 = \frac{1}{n}\sum_{i=1}^{n}\left(x_i - x_0\right)^2.$$

Qu'en est-il de l'écart-type, qu'on souhaite utiliser comme mesure de l'incertitude ? Il est possible de prendre la racine carrée de $\sigma_n^2$

$$\sigma_n = \sqrt{\frac{1}{n-1}\sum_{i=1}^{n}\left(x_i - \overline{x}_n\right)^2}.$$

mais il convient d'être conscient que cet estimateur est biaisé, du fait de la non-linéarité de la fonction racine carrée. Comme elle est concave, cette expression est en réalité une sous-estimation. Il existe des théorèmes et formules approchées pour corriger ce biais, mais cela dépasse le cadre de ce qui est utile à l'agrégation. On se contentera donc de prendre la racine carrée de la variance empirique modifiée.

Il est maintenant légitime de se poser la question de l'incertitude sur ces estimations, en particulier sur celle de la moyenne.

### 3.1.3 Ecart-type de la moyenne

En répétant de nombreuses fois l'expérience consistant à mesurer $n$ valeurs de la grandeur $X$ dont on prend ensuite la valeur moyenne empirique $\overline{x}_n$, on obtient la distribution de probabilité de $\overline{x}_n$. La valeur moyenne de cette distribution est $x_0$, comme déjà démontré. Son écart-type, noté $\sigma_{\overline{x}_n}$ et aussi appelé écart-type de la moyenne, est donné par :

$$\sigma_{\overline{x}_n} = \frac{\sigma}{\sqrt{n}}$$

Cela se démontre facilement puisque

$$\sigma^2_{\overline{x}_n} = \mathsf{E}\left[\overline{X}_n^2\right] - \left(\mathsf{E}\left[\overline{X}_n\right]\right)^2 = \frac{\mathsf{E}\left[X^2\right]}{n} + \frac{n-1}{n}\mathsf{E}\left[X\right]^2 - \mathsf{E}\left[X\right]^2 = \frac{\sigma^2}{n}.$$

$\sigma_{\overline{x}_n}$ est donc l'écart-type de la détermination de $x_0$ à partir de la moyenne de $n$ mesures. Cette détermination est donc $\sqrt{n}$ fois plus précise que celle obtenue à partir d'une mesure unique (voir figure 9). Dans la pratique, l'incertitude décroît lentement (en $1/\sqrt{n}$) et améliorer la précision d'un facteur 10 oblige à effectuer 100 fois plus de mesures.

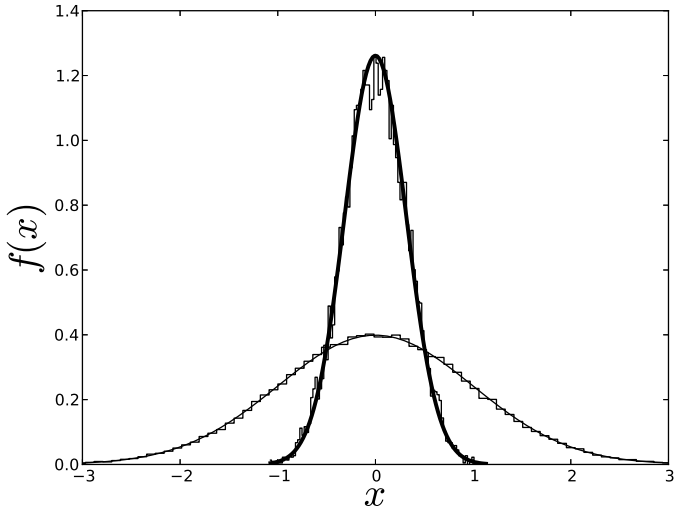


FIGURE 9 – Lorsqu'on effectue une mesure unique d'une grandeur physique $X$ suivant une distribution Gaussienne de moyenne nulle et de variance unité, la valeur trouvée suit la distribution de probabilité représentée en traits fins (l'histogramme est celui de $10^5$ mesures issues de cette distribution). Si l'on moyenne $n$ mesures, les écarts à la moyenne de chacune de ces mesures se compensent statistiquement, avec d'autant plus d'efficacité que $n$ est grand. Si, à partir des $10^5$ mesures précédentes, on réalise $10^4$ déterminations - indépendantes - de la moyenne $\overline{x}_{10}$ de 10 mesures, ces déterminations suivent la distribution représentée par l'histogramme plus piqué autour de zéro. La courbe en traits épais est la distribution Gaussienne de moyenne nulle et de variance 1/10.

En pratique, $\sigma$ n'est pas connu et on utilise son évaluation $\sigma_n$ (l'écart-type empirique modifié, voir plus haut) pour déterminer $\sigma_{\overline{x}_n}$, qu'on estimera donc via $\sigma_n/\sqrt{n}$. Dans le cas où l'on effectue une moyenne sur un petit nombre $n$ de mesures, $\sigma_n$ n'est cependant pas une bonne estimation de $\sigma$ et la théorie montre qu'il faut alors utiliser un coefficient multiplicatif $t$, appelé *coefficient de Student*, et utiliser comme estimateur de $\sigma_{\overline{x}_n}$ l'expression

$$\widehat{\sigma_{\overline{x}_n}} = t\frac{\sigma_n}{\sqrt{n}}.$$

Ce coefficient $t$, qui est tabulé, dépend du nombre de points $n$ moyennés et du niveau de confiance $\mathsf{P}_0$ considéré. À l'agrégation, il est suffisant de ne pas parler de ce coefficient, ce qui revient à le prendre égal à 1 quel que soit le nombre de mesures (voir le tableau 5).

En résumé, si l'on réalise $n$ mesures de $X$, avec les résultats $x_1$, $x_2$,... $x_n$, on écrira le résultat final [18] sous la forme :

$$X = \overline{x}_n \pm t\frac{\sigma_n}{\sqrt{n}}$$

où

$$\overline{x}_n = \frac{x_1 + x_2 + \ldots + x_n}{n} = \frac{1}{n}\sum_{i=1}^{n} x_i$$

et

$$t\frac{\sigma_n}{\sqrt{n}} = t\sqrt{\frac{1}{n(n-1)}\sum_{i=1}^{n}\left(x_i - \overline{x}_n\right)^2}$$

sont les meilleures estimations de la valeur vraie et de l'incertitude-type, respectivement.

## 3.2 Propagation des incertitudes

Dans de nombreux cas, la grandeur physique recherchée n'est pas mesurée directement, mais déduite des mesures d'autres grandeurs via une relation mathématique. On s'intéresse donc ici au problème suivant : connaissant les valeurs des incertitudes $\delta x_1$, ..., $\delta x_n$ sur les grandeurs expérimentales $X_1$, ..., $X_n$, quelle est l'incertitude $\delta y$ sur la grandeur $Y = \phi(X_1, \ldots, X_n)$ ?

### 3.2.1 Cas général

On note $\mu_i$ et $\sigma_i^2$ les moyennes et variances des grandeurs $X_i$. Si l'on peut supposer que ces variables sont indépendantes [19], et que leurs incertitudes sont suffisamment faibles, on peut linéariser $\phi$ au voisinage de la moyenne $(\mu_1, \ldots, \mu_n)$

$$Y = \phi\left(\mu_1, \ldots, \mu_n\right) + \sum_{i=1}^{n}\left(X_i - \mu_i\right)\left(\frac{\partial\phi}{\partial x_i}\right)_{\mu_1,\ldots,\mu_n}$$

de sorte que

$$\mathsf{E}\left[Y\right] = \phi\left(\mu_1, \ldots, \mu_n\right)$$

comme on s'y attendait. Quant à la variance de $Y$, elle s'écrit

$$\sigma_Y^2 = \mathsf{E}\left[Y^2\right] - \mathsf{E}\left[Y\right]^2 = \sum_{i=1}^{n}\sum_{j=1}^{n}\left(\frac{\partial\phi}{\partial x_i}\right)\left(\frac{\partial\phi}{\partial x_j}\right)\mathsf{E}\left[\left(X_i - \mu_i\right)\left(X_j - \mu_j\right)\right]$$

En utilisant l'indépendance des variables $X_i$, les termes pour lesquels $i \neq j$ sont nuls, et on a

$$\sigma_Y^2 = \sum_{i=1}^{n}\left(\frac{\partial\phi}{\partial x_i}\right)^2\sigma_i^2.$$

On en déduit que l'incertitude $\delta y$ se met sous la forme

$$\delta y = \sqrt{\sum_{i=1}^{n}\left(\frac{\partial\phi}{\partial x_i}\right)^2\delta x_i^2}.$$

18. La notation "±" est parfois critiquée, en raison du fait qu'elle sous-entend peut-être que la valeur ne saurait être trouvée en dehors de cet intervalle. Il convient donc de préciser, comme il est indiqué dans [6] que "ce qui suit le ± est l'incertitude-type."

19. On peut ainsi en faire varier une en gardant les autres fixes.

On dit que les incertitudes s'ajoutent quadratiquement.

*Note :* On retrouve ainsi la formule $\sigma_{\overline{x}_n} = \sigma/\sqrt{n}$ reliant l'écart-type de la valeur moyenne de $n$ mesures indépendantes de la même variable à l'écart-type de la variable, puisque dans ce cas :

$$\sigma_{\overline{x}_n} = \sqrt{\sum_{i=1}^{n} \left(\frac{1}{n}\right)^2 \sigma^2} = \frac{\sigma}{\sqrt{n}}$$

*Note :* Un exemple simple pour illustrer cette propriété : si l'on a besoin d'une résistance de 1 kΩ il vaut mieux utiliser 10 résistances de 100 Ω à 1% en série plutôt qu'une seule résistance de 1000 Ω à 1%. En effet, on a dans ce second cas $R = 1000 \pm 10\,\Omega$, alors que dans le premier, $R$ est la somme de $n = 10$ résistances $r = 100\,\Omega$ dont les incertitudes sont $\delta r = 1\,\Omega$, ce qui implique que l'incertitude sur la somme des résistances est

$$\delta R = \sqrt{\sum_{i=1}^{n} \delta r^2} = \sqrt{n} = \sqrt{10}\,\Omega$$

et on a donc $R = 1000 \pm \sqrt{10}\,\Omega$. En pratique, il peut donc être intéressant d'utiliser une boîte à décades plutôt qu'une résistance unique.

### 3.2.2 Comparaison avec l'addition en module

L'addition quadratique donne un résultat plus faible que l'addition en module [20]. Prenant l'exemple de la somme $Y = X_1 + \ldots + X_n$,

$$\delta y = \sqrt{\sum_{i=1}^{n} \delta x_i^2} \leqslant \sum_{i=1}^{n} \delta x_i$$

On peut le comprendre physiquement : la probabilité pour que $x_1 + x_2$ s'écarte de sa valeur moyenne de plus d'un écart type est plus faible que la somme des probabilités correspondantes pour $x_1$ et $x_2$ séparément. En effet, $x_1$ et $x_2$ peuvent s'écarter simultanément de leur valeur moyenne de plus d'un écart-type sans qu'il en soit de même pour leur somme puisque les erreurs peuvent se compenser. En revanche, la somme ne peut s'écarter de sa valeur moyenne de plus d'un écart-type que si l'une au moins des deux variables s'écarte elle aussi fortement de sa valeur moyenne.

### 3.2.3 Méthodologie pratique

En résumé, en présence d'une formule $Y = \phi(X_1, \ldots, X_n)$, on commence par prendre la différentielle des deux membres, on remplace d par $\delta$, l'addition par une addition quadratique et on prend la racine carrée. Par exemple :

$$z = \frac{2}{x} + \frac{3}{y^2} \quad \Longrightarrow \quad \mathrm{d}z = -2\frac{\mathrm{d}x}{x^2} - 6\frac{\mathrm{d}y}{y^3} \quad \Longrightarrow \quad \delta z = \sqrt{\frac{4}{x^4}(\delta x)^2 + \frac{36}{y^6}(\delta y)^2}.$$

20. Dans les anciennes approches des incertitudes, on "majorait l'erreur faite", et on utilisait donc cette addition en module. Il faut l'éviter maintenant.

Pour prendre un exemple plus physique, on considère la force exercée sur une charge $q$ par un champ électrique extérieur $E$ combiné à celui d'une charge ponctuelle $q_0$, en supposant que seuls $E$ et la distance $r$ entre $q$ et $q_0$ sont sujets à incertitudes :

$$F = q\left(E + \frac{q_0}{4\pi\epsilon_0 r^2}\right),$$

ce qui donne

$$\mathrm{d}F = q\left(\mathrm{d}E - \frac{q_0}{2\pi\epsilon_0 r^3}\mathrm{d}r\right),$$

puis

$$\delta F = q\sqrt{(\delta E)^2 + \left(\frac{q_0}{2\pi\epsilon_0}\right)^2 \frac{(\delta r)^2}{r^6}}.$$

### 3.2.4 Calcul direct de l'incertitude relative

Lorsque la fonction $\phi$ contient des produits, rapports ou puissances, on a intérêt à utiliser la différentielle logarithmique pour exprimer directement l'incertitude relative sur $Y$ en fonction des incertitudes sur $X_1$, …, $X_n$. Partant de la forme

$$Y = A\prod_{i=1}^{n} X_i^{\alpha_i},$$

on a

$$\ln Y = \ln A + \sum_{i=1}^{n} \alpha_i \ln X_i \quad \Longrightarrow \quad \frac{\mathrm{d}Y}{Y} = \sum_{i=1}^{n} \alpha_i \frac{\mathrm{d}X_i}{X_i}$$

et on en déduit l'incertitude relative sur $Y$ sous la forme

$$\frac{\delta y}{|y|} = \sqrt{\sum_{i=1}^{n} \alpha_i^2 \left(\frac{\delta x_i}{x_i}\right)^2}$$

▶ *Premier exemple :* La puissance dissipée dans une résistance $R$ aux bornes de laquelle est appliquée une tension $U$ est donnée par $P = U^2/R$. Les valeurs de la résistance et de la tension sont affectées d'incertitudes $\delta R$ et $\delta U$, respectivement. On calcule alors la différentielle logarithmique

$$\frac{\mathrm{d}P}{P} = 2\frac{\mathrm{d}U}{U} - \frac{\mathrm{d}R}{R}$$

d'où l'on déduit que

$$\frac{\delta P}{P} = \sqrt{4\left(\frac{\delta U}{U}\right)^2 + \left(\frac{\delta R}{R}\right)^2}$$

▶ *Deuxième exemple :* La vitesse de précession $\Omega$ d'un gyroscope déséquilibré est donnée par

$$\Omega = \frac{mga}{J\omega}$$

en fonction de la vitesse de rotation $\omega$, du moment d'inertie $J$, de la masse $m$ déséquilibrant le gyroscope et de la distance $a$ entre la position de cette masse $m$ et le centre de masse du

gyroscope équilibré. En supposant que l'accélération de la pesanteur $g$ est parfaitement connue, on a la différentielle logarithmique

$$\frac{\mathrm{d}\Omega}{\Omega} = \frac{\mathrm{d}m}{m} + \frac{\mathrm{d}a}{a} - \frac{\mathrm{d}J}{J} - \frac{\mathrm{d}\omega}{\omega}$$

et donc

$$\frac{\delta\Omega}{\Omega} = \sqrt{\left(\frac{\delta m}{m}\right)^2 + \left(\frac{\delta a}{a}\right)^2 + \left(\frac{\delta J}{J}\right)^2 + \left(\frac{\delta \omega}{\omega}\right)^2}$$

▶ *Troisième exemple :* Le taux d'ondes stationnaires (TOS) dans un guide d'onde est défini à partir des amplitudes (positives) du champ électrique incident $E_i$ et réfléchi $E_r$ comme

$$t = \frac{E_i + E_r}{E_i - E_r}$$

On suppose que $E_i > E_r$, de sorte que dénominateur comme numérateur sont des quantités positives. Cet exemple est intéressant car les influences de $E_i$ au numérateur et au dénominateur se compensent partiellement (si $E_i$ augmente, numérateur et dénominateur augmentent donc $t$ varie peu). La différentielle logarithmique conduit à :

$$\frac{\mathrm{d}t}{t} = \frac{\mathrm{d}E_i + \mathrm{d}E_r}{E_i + E_r} - \frac{\mathrm{d}E_i - \mathrm{d}E_r}{E_i - E_r} = \left[\frac{1}{E_i + E_r} - \frac{1}{E_i - E_r}\right]\mathrm{d}E_i + \left[\frac{1}{E_i + E_r} + \frac{1}{E_i - E_r}\right]\mathrm{d}E_r$$

La formule de propagation des incertitudes n'étant valable que pour des variables indépendantes, il faut impérativement séparer les contributions liées à $\mathrm{d}E_i$ et $\mathrm{d}E_r$, soit

$$\frac{\mathrm{d}t}{t} = \frac{2E_i}{E_i^2 - E_r^2}\mathrm{d}E_r - \frac{2E_r}{E_i^2 - E_r^2}\mathrm{d}E_i$$

et donc

$$\frac{\delta t}{t} = \sqrt{\frac{4E_i^2}{\left(E_i^2 - E_r^2\right)^2}\left(\delta E_r\right)^2 + \frac{4E_r^2}{\left(E_i^2 - E_r^2\right)^2}\left(\delta E_i\right)^2} = \frac{2}{|E_i^2 - E_r^2|}\sqrt{E_i^2\left(\delta E_r\right)^2 + E_r^2\left(\delta E_i\right)^2}.$$

On note qu'en particulier si $E_i \gg E_r$, la compensation évoquée ci-dessus se traduit par la faiblesse du coefficient devant $\left(\delta E_i\right)^2$.

### 3.2.5 Approche "Monte Carlo"

Le calcul ci-dessus montre à quel point la détermination des incertitudes peut devenir ardue analytiquement, lorsque les grandeurs d'intérêt dépendent de manière complexe des mesures faites en pratique ou encore des paramètres environnementaux. Pour ces derniers, on ne connaît d'ailleurs pas toujours leur influence, en tous cas pas analytiquement.

Dans ce cas, on peut avoir recours à des méthodes dites *Monte Carlo*, en référence aux casinos de la principauté. Connaissant les distributions de probabilité des grandeurs d'influence (dans l'exemple ci-dessus, $E_i$ et $E_r$), on peut en faire des tirages aléatoires, calculer pour chacun de ces tirages la grandeur d'intérêt, (ici $t$) et en déduire petit à petit la distribution de probabilité de $t$, et en tirer sa moyenne et son écart-type. Le logiciel "GUM_MC" est un outil efficace pour calculer des incertitudes composées sur ce principe [21].

21. `http://jeanmarie.biansan.free.fr/gum_mc.html`

# 4 Quelques remarques utiles et exemples pratiques

## 4.1 Présentation d'un résultat expérimental, chiffres significatifs

L'écriture rapportant la mesure d'une grandeur physique $X$ est $X = x \pm \delta x$, où $x$ est la meilleure estimation de la valeur vraie $x_0$ et $\delta x$ l'incertitude-type sur la mesure (incertitude absolue). On peut aussi utiliser l'incertitude-type relative ou fractionnaire : $\delta x/|x|$ qu'on exprime la plupart du temps en pourcentages.

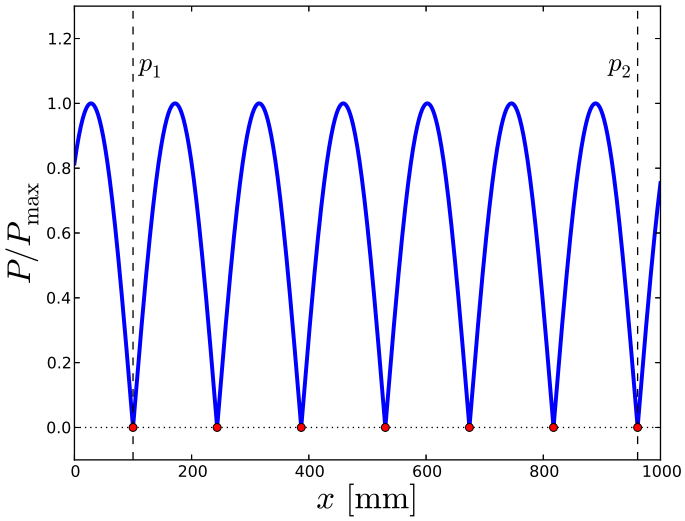


FIGURE 10 – Pression acoustique en fonction de la position d'un micro dans une expérience d'ondes stationnaires. Les points repèrent les nœuds de pression. L'expérimentateur mesure la position de deux d'entre eux, $p_1$ et $p_2$.

Cette incertitude - au sens commun du terme - doit se refléter dans le nombre de chiffres significatifs utilisés pour présenter le résultat. Par exemple, dans une expérience d'ondes stationnaires en acoustique (figure 10) on mesure les positions $p_1$ et $p_2$ de deux nœuds de pression entre lesquels se trouvent 5 autres noeuds, de sorte que la distance $p_2 - p_1$ correspond à trois longueurs d'onde $\lambda$. On trouve ainsi $p_2 - p_1 = 3\lambda = 862\,\text{mm}$, avec une incertitude-type absolue estimée à 10 mm pour chaque mesure de position (parce qu'on a du mal à repérer précisément les nœuds). À la calculatrice on obtient pour la longueur d'onde et pour l'incertitude-type absolue les résultats suivants [22] :

$$\overline{\lambda} = \frac{862}{3} = 287,33333333333\,\text{mm} \qquad \text{et} \qquad \delta\lambda = \frac{1}{3}\sqrt{10^2 + 10^2} = 4,71404520791\,\text{mm}.$$

L'incertitude-type étant toujours évaluée grossièrement, on garde 1 à 2 chiffres significatifs [23] :

$$\delta\lambda = 5\ \text{mm}.$$

22. On note la division par $n = 3$ et non par $\sqrt{n}$ (voir 3.1.3) car ici on mesure une longueur égale à $n\lambda$ et il ne s'agit pas de mesurer $n$ fois une longueur $\lambda$. On remarque qu'il est donc plus avantageux de mesurer une fois une longueur de $n\lambda$ que $n$ fois une longueur de $\lambda$.

23. On en garde généralement 2, pour éviter de trop grossières erreurs d'arrondi. Mais ici on n'en garde qu'un...

Le dernier chiffre significatif de la mesure doit être cohérent avec l'incertitude. On écrira donc :

$$\lambda = 287 \pm 5 \text{ mm} = 287 \text{ mm} \pm 2\%.$$

## 4.2 Comparaison entre valeur mesurée et valeur de référence

Ayant obtenu la valeur mesurée $x$ avec son incertitude-type $\delta x$, on la compare à la valeur de référence (pour une valeur expérimentale de référence, ne pas parler de valeur exacte, mais de *valeur tabulée*) $x_t$. Il n'est pas anormal que l'intervalle $[x - \delta x, x + \delta x]$ ne contienne pas la valeur de référence. Ainsi dans le cas fréquent d'une distribution Gaussienne, le tableau 3 montre qu'il y a 32 % de chances - soit environ une chance sur trois - d'être dans ce cas. On commencera à douter de la mesure lorsque l'écart $\varepsilon = |x - x_t|$ atteint plus de 2 écarts-type (probabilité que la mesure soit bonne : 1 chance sur 20 à $2\delta x$, 1 chance sur 100 à $2{,}5\delta x$, et 1 chance sur $10^4$ à $4\delta x$ pour une distribution Gaussienne). Si c'est le cas il faut chercher de possibles sources d'erreurs systématiques.

Par exemple, on mesure la vitesse du son dans l'air à 20°C. On effectue une série de mesures qui conduit à $v = 335$ m.s$^{-1}$ et à une incertitude-type $\delta v = 5$ m.s$^{-1}$. On a donc

$$v = 335 \pm 5 \text{ m.s}^{-1}$$

La valeur tabulée indique à cette température : $v_t = 343$ m.s$^{-1}$. Elle est en-dehors du domaine d'incertitude, mais l'écart entre 343 et 335 vaut $1{,}6\delta v$. La table 3 indique qu'il y a environ une chance sur 10 pour que ce cas se produise. On pourra considérer que la mesure est valide, mais néanmoins pas très satisfaisante. On pourra aussi s'interroger sur la validité de la mesure de température.

Cette approche, appelée *z-score* dans les programmes actuels [6], définit la grandeur

$$z = \frac{|x - x_t|}{\delta x}$$

qui correspond donc à la variable normalisée de la figure 5. On pourra considérer qu'un $z$-score inférieur ou égal à deux correspond à une "bonne mesure", tandis qu'un $z$-score supérieur à deux indique un problème (une chance sur 20 que cela se produise pour une distribution Gaussienne). Il faut cependant garder à l'esprit que ce critère est assez arbitraire. Rappelons aussi que les erreurs systématiques n'apparaissent pas dans un calcul d'incertitude, c'est pourquoi on les supposera négligeables, quitte à revenir à la fin sur cette supposition en cas de désaccord entre la valeur trouvée et la valeur de référence, c'est-à-dire si le $z$-score est supérieur à 2.

## 4.3 Analyse statistique d'une série de mesures

Huit étudiants mesurent la longueur d'onde de la raie verte du mercure en utilisant une fente fine éclairée par la lampe, une lentille et un réseau. Ils obtiennent les résultats suivants :

TABLE 4 – Mesures indépendantes de la raie verte du mecure

| $i$ (n° de l'étudiant) | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 |
|---|---|---|---|---|---|---|---|---|
| $\lambda_i$ (nm) | 538,2 | 554,3 | 545,7 | 552,3 | 566,4 | 537,9 | 549,2 | 540,3 |

En utilisant un logiciel de traitement des données [24], on obtient la moyenne empirique et l'écart-type empirique modifié :

$$\overline{\lambda}_n = 548,04 \text{ nm} \qquad \text{et} \qquad \sigma_n = 9,72 \text{ nm}$$

On en déduit que l'incertitude-type sur la moyenne empirique des 8 mesures vaut :

$$\delta\lambda = \sigma_{\overline{\lambda}_n} = 9,72/\sqrt{8} = 3,44 \text{ nm}$$

On peut donc écrire finalement :

$$\lambda = 548 \pm 3 \text{ nm} \qquad \text{ou} \qquad 545 \text{ nm} \leqslant \lambda \leqslant 551 \text{ nm}$$

Remarquer qu'on arrondit 3,44 par 3 qui est sa plus proche valeur avec un seul chiffre significatif [25]. Dans les anciens calculs d'incertitudes, où l'incertitude était un majorant de l'erreur, on aurait arrondi à la valeur supérieure. Notons aussi que lorsqu'on calcule une incertitude composée, c'est cette dernière qu'on arrondit à la dernière étape, car arrondir les incertitudes intermédiaires risque de provoquer de grossières erreurs, et de mal analyser le $z$-score.

On peut maintenant comparer à la valeur tabulée : $\lambda_t = 545,07$ nm et conclure qu'il y a une bonne concordance à $1\,\sigma$.

*Note :* L'incertitude-type $\delta\lambda$ est calculée à partir de $\sigma_n$ qui est une estimation de l'écart-type de la distribution des mesures de $\lambda$. Cette estimation, correcte dans le cas d'un grand nombre de mesures, devient imprécise si ce nombre est faible. La loi de Student mentionnée au 3.1.3 permet de tenir compte de cet effet. Le tableau 5 donne les valeurs du coefficient $t$ en fonction du nombre de mesures pour un intervalle de confiance de 68%.

TABLE 5 – Coefficient de Student pour un niveau de confiance $\mathsf{P}_0 = 68\%$.

| $n$ | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 20 | 40 | $\infty$ |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| $t_{n;68\%}$ | 1,84 | 1,32 | 1,20 | 1,14 | 1,11 | 1,09 | 1,08 | 1,07 | 1,06 | 1,03 | 1,01 | 1,00 |

Pour 8 mesures, on a $t_{8;68\%} = 1,08$. On doit donc écrire en toute rigueur :

$$\lambda = 548,04 \pm 1,08 \times 9,72/\sqrt{8} = 548 \pm 4 \text{ nm}$$

Cet exemple montre qu'on peut oublier la correction de Student dans le cadre des TP usuels.

## 4.4 Incertitude relative, termes dominants

On détermine une puissance électrique par la formule $P = RI^2$. On mesure $R = 15,7\,\Omega$ avec $\delta R = 1\,\Omega$ (à cause de problèmes de contact) et $I = 0,274$ A avec $\delta I = 0,002$ A (d'après le fabricant de l'appareil). On en déduit la formule de propagation des incertitudes indépendantes :

$$\frac{\delta P}{P} = \sqrt{\left(\frac{\delta R}{R}\right)^2 + 4\left(\frac{\delta I}{I}\right)^2} = \sqrt{4.10^{-3} + 2.10^{-4}} \simeq \sqrt{4.10^{-3}} \simeq 0,06 = 6\%$$

24. En pratique :
- ▶ Sous `Qtiplot`, utiliser `Analyse` → `Statistiques descriptives`
- ▶ Sous `Igor Pro`, utiliser `Analysis` → `Statistics` : $\mathrm{V_{avg}}$ donne $\overline{x}_n$ et $\mathrm{V_{sdev}}$ donne $\sigma_n$
- ▶ Sous `Synchronie`, le module `Statistiques` donne $\sigma_n\sqrt{(n-1)/n}$ et non pas $\sigma_n$.
- ▶ Sous `Excel`, utiliser `Insertion` → `Fonction` : `MOYENNE` donne $\overline{x}_n$ et `ECARTYPE` donne $\sigma_n$.
- ▶ Sous `Kaleidagraph`, utiliser `Functions` → `Statistics` : `Mean` donne $\overline{x}_n$ et `Std Deviation` donne $\sigma_n$.

25. Dans le cas où un 5 est le dernier chiffre, on arrondit par excès.

On remarque que l'incertitude sur l'intensité est négligeable. De manière générale, il est important de comparer les incertitudes des différentes grandeurs mesurées. Ici,

$$\frac{\delta I}{I} \simeq 1\% \quad \text{et} \quad \frac{\delta R}{R} \simeq 6\%.$$

Il est clair que dans l'exemple ci-dessus, la principale cause d'erreur porte sur la résistance et que si l'on veut améliorer la mesure de $P$, c'est sur elle qu'il faut porter ses efforts. Il ne servirait à rien d'acheter un ampèremètre haut de gamme pour diminuer $\delta I$, qui est déjà négligeable.

## 4.5 Petits facteurs

Pour trouver la capacité thermique massique $c$ d'un liquide lors d'une expérience de calorimétrie, on mesure l'élévation de température $\theta_2 - \theta_1$ d'une certaine masse $m$ de ce liquide, dans laquelle on dissipe pendant un temps $t$ une énergie électrique connue, via une résistance chauffante plongée dans le liquide, alimentée sous une tension $U$ et parcourue par un courant $I$. On obtient la capacité thermique cherchée par la formule :

$$c = \frac{UIt}{m(\theta_2 - \theta_1)}.$$

On a mesuré :

$$t = 153\,\mathrm{s} \qquad m = 345\,\mathrm{g} \qquad \theta_1 = 19,3\,^\circ\mathrm{C} \qquad \theta_2 = 20,3\,^\circ\mathrm{C}$$

avec les incertitudes suivantes :

$$\frac{\delta U}{U} = 1\% \qquad \frac{\delta I}{I} = 2\% \qquad \delta t = 2\,\mathrm{s} \qquad \delta m = 5\,\mathrm{g} \qquad \delta\theta_1 = \delta\theta_2 = \delta\theta = 0{,}1\,^\circ\mathrm{C}$$

Le calcul d'incertitude montre que la seule incertitude qui joue un rôle est celle sur les températures, car

$$\frac{\delta U}{U} = 1\% \qquad \frac{\delta I}{I} = 2\% \qquad \frac{\delta t}{t} = 1\% \qquad \frac{\delta m}{m} = 1\% \qquad \frac{\delta\theta}{\theta_2 - \theta_1} = 10\%$$

On obtient finalement :

$$\frac{\delta c}{c} = \sqrt{\left(\frac{\delta U}{U}\right)^2 + \left(\frac{\delta I}{I}\right)^2 + \left(\frac{\delta t}{t}\right)^2 + \left(\frac{\delta m}{m}\right)^2 + 2\left(\frac{\delta\theta}{|\theta_2 - \theta_1|}\right)^2} \simeq \frac{\sqrt{2}\delta\theta}{|\theta_2 - \theta_1|} = 14\%$$

On constate avec cet exemple qu'il faut si possible éviter l'apparition de petites différences : l'incertitude relative sur la différence de 2 nombres très voisins est grande, alors que l'incertitude relative sur chaque mesure est faible :

$$\frac{\delta\theta}{\theta_1} = \frac{\delta\theta}{\theta_2} = 0{,}03\%$$

Pour améliorer cette expérience il faudrait accroître nettement la variation de température en augmentant la durée de l'expérience, l'intensité du courant, ou les deux.

## 4.6 Indications des appareils : évaluation de type B

Si l'on ne dispose pas du temps nécessaire pour faire une série de mesures, on procède à une évaluation de type B, qui consiste à estimer l'incertitude $\delta x$ à partir des spécifications des appareils de mesure et des conditions expérimentales. Par exemple, en ouvrant la notice d'un voltmètre numérique, on trouve typiquement les indications suivantes :

*"La précision* $\Delta$ *de la mesure est donnée par* $\Delta = \pm 2$ *fois le dernier digit* $\pm 0,1\%$ *de la valeur lue."*

On peut considérer que l'indication donnée par le fabricant a deux origines :
► celle qui provient d'une erreur d'étalonnage (variable d'un appareil à l'autre et d'un calibre à l'autre sur un même appareil). Cette erreur est une erreur systématique quand on utilise le même calibre d'un même appareil. Elle devient aléatoire quand on utilise plusieurs calibres ou plusieurs appareils, même de modèle identique. Elle est essentiellement présente dans les ± 0,1% de la valeur lue.
► celle qui provient d'erreurs aléatoires (bruit de mesure). Elle est essentiellement présente dans les ± 2 fois le dernier digit.

À titre d'exemple [6], on considère une notice de voltmètre indiquant, sur le calibre 20 V, une précision "$\pm(0.5\% + 8)$" et une résolution de 0.01 V. Lorsque l'appareil affiche 10 V, la précision à indiquer est

$$\Delta = 0.5\% \times 10 + 8 \times 0.01 = 0.05 + 0.08 = 0.13\,\mathrm{V}$$

Quel est le lien entre cette *précision* $\Delta$ et l'incertitude-type $\delta x$ ? Deux cas se présentent :
► Si le constructeur donne un niveau de confiance, en précisant par exemple que $\Delta$ correspond à 3 écart-types, on en déduit facilement l'incertitude-type $\delta x = \Delta/3$. Dans ce cas, le fabricant s'engage sur la précision de son appareil.
► Dans le cas contraire, la précision donnée n'est qu'indicative. Il est alors usuel de supposer une distribution de probabilité uniforme de largeur $2\Delta$, ce qui conduit par un calcul simple (voir 2.8.1) à l'incertitude-type $\delta x = \Delta/\sqrt{3}$. C'est le cas de la plupart des appareils de mesure courants. On peut aussi supposer une distribution triangulaire, ce qui implique la relation $\delta x = \Delta/\sqrt{6}$, intermédiaire entre les cas rectangulaire et Gaussien.

## 4.7 Niveaux de confiance variables

Certains fabricants indiquent la précision de leurs appareils en utilisant l'incertitude-type qui correspond à 1 écart-type donc à un niveau de confiance de 68%. D'autres utilisent l'incertitude-élargie qui correspond à 2 écart-types soit un niveau de confiance d'environ 95%. Il faut donc prendre garde à quelle incertitude fait référence le constructeur.

Par exemple, pour déterminer une puissance par la relation $P = UI$, on mesure $I = 1,27\,\mathrm{A}$ avec une incertitude-type égale à $0,02\,\mathrm{A}$, et on mesure $U = 15,24\,\mathrm{V}$ avec une incertitude-élargie égale à $0,06\,\mathrm{V}$. On peut donc poser : $\delta I = 0,02\,\mathrm{A}$ et comme l'incertitude-élargie vaut le double de l'incertitude-type : $\delta U = 0,03\,\mathrm{V}$. On en déduit l'incertitude-type relative sur $P$ :

$$\frac{\delta P}{P} = \sqrt{\left(\frac{0,02}{1,27}\right)^2 + \left(\frac{0,03}{15,24}\right)^2} \simeq 2\%.$$

Notons qu'avec les incertitudes choisies ici, celle sur $I$ domine celle sur $U$, de sorte que cette subtilité est superflue dans ce cas précis.

## 4.8 Erreurs aléatoires liées

Comme on l'a déjà dit, la loi de propagation des incertitudes écrite plus haut suppose que les variables mises en jeu ont des erreurs aléatoires indépendantes. Si cette condition n'est pas vérifiée, la formule de combinaison quadratique ne s'applique pas. Par contre, l'inégalité non quadratique suivante est toujours vérifiée :

$$\delta y \leqslant \sum_{i=1}^{n} \left| \frac{\partial \phi}{\partial x_i} \right| \delta x_i$$

Elle s'obtient facilement à partir du développement linéaire de $\phi$ au voisinage de $(\mu_1, \ldots, \mu_n)$. On va voir cependant sur un exemple pratique que cette inégalité risque d'être de peu d'intérêt.

Un fabricant de voltmètres grand public effectue un test statistique sur ses appareils et trouve que l'incertitude-type relative vaut 1%. Il l'indique dans la notice en précisant que l'incertitude aléatoire de chaque appareil est négligeable (si on fait 10 fois la même mesure avec le même appareil on obtient le même résultat) et que l'incertitude annoncée provient des défauts d'étalonnage d'un appareil à l'autre (par mesure d'économie les appareils ne sont pas réglés en sortie de la chaîne de production).

L'utilisateur mesure 2 tensions voisines $V_1 = 12,71$ V et $V_2 = 9,32$ V afin de déterminer leur différence $V_3 = V_1 - V_2$ et leur somme $V_4 = V_1 + V_2$.

▶ S'il utilise deux voltmètres, l'un pour mesurer $V_1$, l'autre pour mesurer $V_2$, les incertitudes $\delta V_1 = 12,71 \times 0,01 = 0,13$ V et $\delta V_2 = 9,32 \times 0,01 = 0,09$ V sont indépendantes et il applique donc la combinaison quadratique :

$$\delta V_3 = \delta V_4 = \sqrt{(0,13)^2 + (0,09)^2} = 0,16 \text{ V}.$$

▶ S'il utilise le même voltmètre sur le même calibre pour mesurer $V_1$ et $V_2$, les deux incertitudes deviennent liées (si $V_1$ est trop grand, $V_2$ est trop grand en proportion). En utilisant la notation différentielle, on peut écrire : $\mathrm{d}V_3 = \mathrm{d}V_1 - \mathrm{d}V_2$ avec $\mathrm{d}V_2/V_2 = \mathrm{d}V_1/V_1$ puisqu'il y a uniquement une erreur d'étalonnage. On en déduit

$$\mathrm{d}V_3 = \mathrm{d}V_1 \left(1 - \frac{V_2}{V_1}\right)$$

puis, en passant aux incertitudes,

$$\delta V_3 \lesssim \delta V_1 \left|1 - \frac{V_2}{V_1}\right| = 0,035 \text{ V}$$

qui est très inférieur à la valeur obtenue avec 2 voltmètres car on est sûr que les erreurs sur $V_1$ et $V_2$ se compensent partiellement. À la limite, si $V_1 = V_2$, l'incertitude sur la différence est nulle car on est certain que, bien que fausses, les mesures de ces deux tensions identiques sont identiquement fausses. De même, on trouve

$$\delta V_4 \lesssim \delta V_1 \left|1 + \frac{V_2}{V_1}\right| = 0,22 \text{ V}.$$

Pour $V_4$, cette majoration de l'incertitude obtenue avec un seul voltmètre est supérieure à l'estimation de l'incertitude de la mesure à deux voltmètres car on est sûr que les erreurs s'additionnent. En réalité chaque appareil a en plus une incertitude aléatoire qui lui est propre et qui vient compliquer l'étude mais atténuer l'effet des erreurs liées.

### 4.9 Analyse des causes d'erreur

Dans une expérience d'interférences avec les fentes d'Young, on mesure la distance $d$ (de l'ordre du mètre) entre la bifente et l'écran avec une règle graduée en centimètres. On estime alors généralement l'incertitude-type $\delta x$ sur la mesure à $1/4$ de graduation. Et si l'on utilise une règle graduée en millimètres ? Décider d'une incertitude-type égale à 0,25 mm serait illusoire : en effet au fur et à mesure que la précision d'affichage de l'instrument augmente, il faut accroître l'analyse des causes d'erreur. Ainsi dans le cas présent la position de la bifente placée dans son support est difficile à repérer précisément à l'échelle du mm alors qu'elle était facile à repérer à l'échelle du cm.

## 5 Vérification d'une loi physique

On cherche souvent à vérifier une loi physique reliant deux grandeurs distinctes. Ainsi on peut chercher à vérifier que la tension de Hall $V_H$ aux bornes d'un échantillon semi-conducteur est reliée linéairement au champ magnétique $B$ par

$$V_H = \frac{IB}{qbn_p}$$

où $I$ est le courant qui traverse l'échantillon, $q$ la charge d'un porteur, $b$ l'épaisseur de l'échantillon dans la direction de $B$ et $n_p$ est le nombre de porteurs par unité de volume du semi-conducteur. En mesurant $V_H$ pour différentes valeurs de $B$ à $I$ fixée, on souhaite d'une part vérifier expérimentalement que la relation liant $V_H$ à $B$ est linéaire et d'autre part déterminer la valeur de $n_p$. Plus généralement, on cherche à vérifier que deux grandeurs $X$ et $Y$ sont reliées par une loi du type $Y = f_\alpha(X)$ où $f_\alpha$ est une fonction dépendant d'un certain nombre de paramètres, notés collectivement $\alpha$, qu'on cherche à déterminer. Dans l'exemple précédent, on a $\alpha = \{n_p\}$. On effectue une série de mesures $y_1$, $y_2$,..., $y_n$ de la grandeur physique $Y$, correspondant à une série de $n$ valeurs $x_1$, $x_2$,..., $x_n$ de l'autre grandeur $X$. L'objet des paragraphes suivants est de montrer comment on détermine les "meilleures" valeurs des paramètres $\alpha$ à partir de ces mesures, d'abord dans le cas linéaire puis dans le cas général. On présente également des outils qui permettent d'estimer dans quelle mesure la fonction $f_\alpha$ ainsi déterminée décrit correctement ou non les données expérimentales.

### 5.1 Régression linéaire $Y = a + bX$

Dans de nombreux cas, on cherche à savoir dans quelle mesure des données expérimentales s'accordent avec une loi affine du type $Y = a + bX$. On cherche également une estimation des paramètres $a$ et $b$ et on souhaite connaître la précision de cette estimation.

On supposera ici que les incertitudes sur $X$ sont négligeables devant celles sur $Y$ (on peut souvent se ramener à cette situation car il est très fréquent que les incertitudes relatives sur une variable soient beaucoup plus faibles que les incertitudes relatives sur l'autre). Il existe une littérature extensive sur les méthodes de régression linéaire dans le cas où il existe des incertitudes à la fois en abscisses et en ordonnées. On pourra consulter par exemple [3].

On dispose donc d'un tableau de $n$ mesures $(x_1, y_1), (x_2, y_2), \ldots, (x_n, y_n)$ et éventuellement, pour chacune de ces mesures, de l'incertitude associée à la mesure de $Y$ (on note alors $\sigma_{i,\mathrm{e}}$ l'incertitude-type sur $y_i$). À titre d'exemple, on utilise les mesures regroupées dans le tableau 6 (ici $n = 12$).

On appelle ensuite le programme de modélisation mathématique (ou *ajustement*, ou encore *fit*, mais on évitera les anglicismes) du logiciel utilisé. Celui-ci trace la "meilleure" droite passant

TABLE 6 – Données utilisées pour les exemples de régression linéaire

| $x_i$ | 0 | 2 | 4 | 6 | 8 | 10 |
|---|---|---|---|---|---|---|
| $y_i$ | 14,79 | 33,52 | 36,50 | 51,88 | 63,11 | 66,94 |
| $x_i$ | 12 | 14 | 16 | 18 | 20 | 22 |
| $y_i$ | 74,58 | 92,46 | 89,50 | 109,29 | 117,40 | 118,37 |

par les points en utilisant les résultats du paragraphe suivant. Les résultats sont présentés sur la figure 11.

### 5.1.1 Meilleure estimation des paramètres $a$ et $b$

Dans le cas général où les mesures de $Y$ ont des incertitudes-types différentes, à savoir qu'on précise pour chaque mesure $y_i$ l'incertitude-type expérimentale $\sigma_{i,\mathrm{e}}$, on introduit le poids de chaque mesure par $w_i = 1/\sigma_{i,\mathrm{e}}^2$. Ce poids est d'autant plus grand que la mesure est fidèle. La méthode utilisée, appelée *méthode des moindres carrés*, consiste à chercher les valeurs de $a$ et $b$ qui rendent minimum l'écart quadratique moyen pondéré

$$g(a,b) = \sum_{i=1}^{n} w_i \left[y_i - (a + bx_i)\right]^2 .$$

En dérivant cette expression par rapport aux paramètres $a$ et $b$, on obtient sans difficulté les estimations optimales des paramètres de la régression linéaire, c'est-à-dire les valeurs $a_0$ et $b_0$ qui annulent ces dérivées partielles. On aboutit à

$$a = a_0 = \frac{\left(\sum_{i=1}^{n} w_i x_i^2\right)\left(\sum_{i=1}^{n} w_i y_i\right) - \left(\sum_{i=1}^{n} w_i x_i\right)\left(\sum_{i=1}^{n} w_i x_i y_i\right)}{\Delta}$$

$$b = b_0 = \frac{\left(\sum_{i=1}^{n} w_i\right)\left(\sum_{i=1}^{n} w_i x_i y_i\right) - \left(\sum_{i=1}^{n} w_i x_i\right)\left(\sum_{i=1}^{n} w_i y_i\right)}{\Delta}$$

où

$$\Delta = \left(\sum_{i=1}^{n} w_i\right)\left(\sum_{i=1}^{n} w_i x_i^2\right) - \left(\sum_{i=1}^{n} w_i x_i\right)^2$$

Dans le cas où toutes les mesures de $Y$ ont la même incertitude-type (ce qui est supposé implicitement si l'on n'a pas précisé les incertitudes-types $\sigma_{i,\mathrm{e}}$) les poids sont tous identiques, $w_i = w$, ils se simplifient et on obtient

$$a = a_0 = \frac{\left(\sum_{i=1}^{n} x_i^2\right)\left(\sum_{i=1}^{n} y_i\right) - \left(\sum_{i=1}^{n} x_i\right)\left(\sum_{i=1}^{n} x_i y_i\right)}{\Delta}$$

$$b = b_0 = \frac{n\left(\sum_{i=1}^{n} x_i y_i\right) - \left(\sum_{i=1}^{n} x_i\right)\left(\sum_{i=1}^{n} y_i\right)}{\Delta}$$

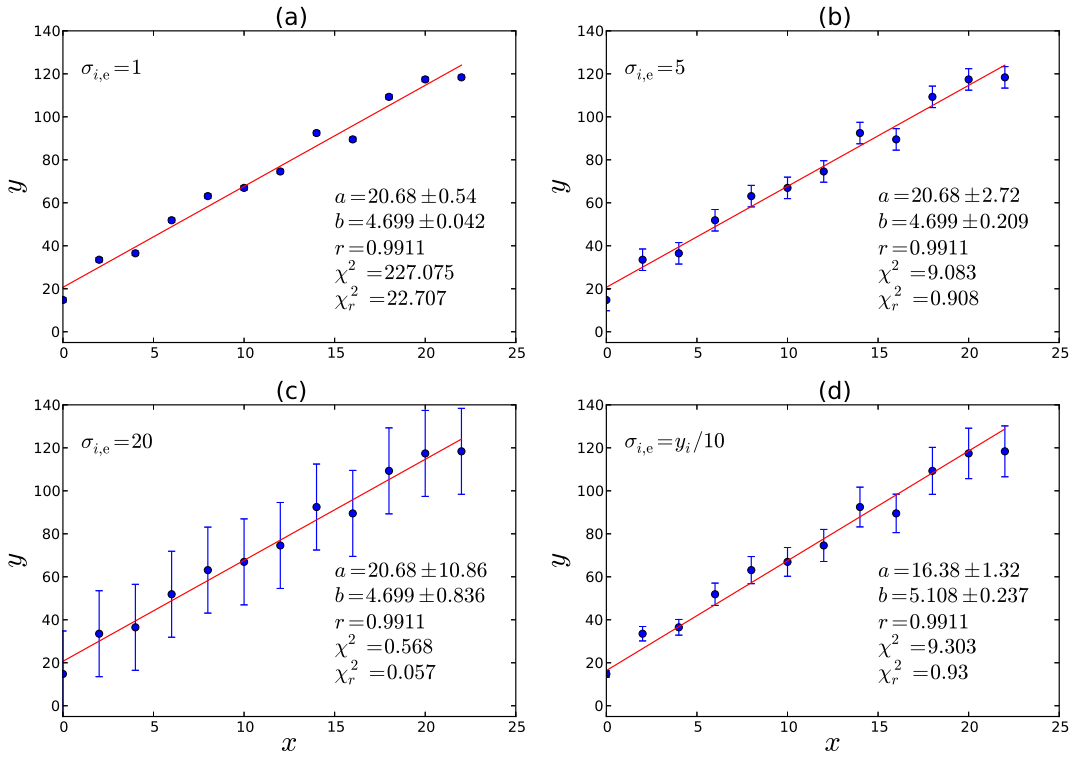


FIGURE 11 – Régressions linéaires $Y = a + bX$ sur l'ensemble des points de mesure du tableau 6, en utilisant comme incertitudes expérimentales sur les mesures $y_i$ les valeurs $\sigma_{i,\mathrm{e}} = 1$ (a), $\sigma_{i,\mathrm{e}} = 5$ (b), $\sigma_{i,\mathrm{e}} = 20$ (c) et $\sigma_{i,\mathrm{e}} = y_i/10$ (d).

où

$$\Delta = n\sum_{i=1}^{n} x_i^2 - \left(\sum_{i=1}^{n} x_i\right)^2$$

*Note :* Ces formules ne sont bien entendues applicables que pour $n \geqslant 2$. D'ailleurs, pour $n = 1$ elles donnent des formes indéterminées pour $a$ et $b$.

La figure 11 montre les résultats obtenus pour les données du tableau 6 dans différents cas de figure. On constate notamment que les valeurs de $a$ et $b$ sont identiques quelle que soit la valeur de $\sigma_{i,\mathrm{e}}$ si cette incertitude est la même pour tous les points de mesure. Dans le cas contraire, les estimations de $a$ et $b$ dépendent des poids.

### 5.1.2 Incertitudes-type $\sigma_a$ et $\sigma_b$ sur les paramètres $a$ et $b$

Le programme d'ajustement fournit également les valeurs des incertitudes-types $\sigma_a$ et $\sigma_b$ sur $a$ et $b$. Pour cela, on utilise les formules donnant les estimations $a_0$ et $b_0$ et les techniques de propagation des incertitudes indépendantes décrites au 3.2. Dans le cas le plus général, cela donne

$$\sigma_a = \sqrt{\frac{\sum_{i=1}^{n} w_i x_i^2}{\Delta}} \qquad \text{et} \qquad \sigma_b = \sqrt{\frac{\sum_{i=1}^{n} w_i}{\Delta}}$$

ce qu'on peut démontrer de la manière suivante (on montre ici le calcul pour $\sigma_b$, celui pour $\sigma_a$ étant parfaitement similaire) en rappelant que $\mathrm{d}x_i = 0$ par hypothèse :

$$\mathrm{d}b = \sum_{i=1}^{n} \frac{\left[\left(\sum_{j=1}^{n} w_j\right) w_i x_i - \left(\sum_{j=1}^{n} w_j x_j\right) w_i\right] \mathrm{d}y_i}{\Delta}$$

d'où l'on tire, sachant que $\sigma_{i,\mathrm{e}}^2 = 1/w_i$,

$$\sigma_b = \sqrt{\sum_{i=1}^{n} \frac{\left[\left(\sum_{j=1}^{n} w_j\right) w_i x_i - \left(\sum_{j=1}^{n} w_j x_j\right) w_i\right]^2}{w_i \Delta^2}}$$

En développant le carré, on a

$$\sum_{i=1}^{n} \frac{\left[\left(\sum_{j=1}^{n} w_j\right) w_i x_i - \left(\sum_{j=1}^{n} w_j x_j\right) w_i\right]^2}{w_i} = \Delta \sum_{i=1}^{n} w_i$$

d'où l'on tire la formule donnant $\sigma_b$.
Si toutes les mesures de $Y$ ont la même incertitude, ce résultat prend la forme suivante

$$\sigma_a = \sigma_y \sqrt{\frac{\sum_{i=1}^{n} x_i^2}{n\sum_{i=1}^{n} x_i^2 - \left(\sum_{i=1}^{n} x_i\right)^2}} \qquad \text{et} \qquad \sigma_b = \sigma_y \sqrt{\frac{n}{n\sum_{i=1}^{n} x_i^2 - \left(\sum_{i=1}^{n} x_i\right)^2}}$$

où $\sigma_y$ est une incertitude sur $Y$ qui dépend du cas de figure :
▶ Si l'on a fourni l'incertitude-type expérimentale $\sigma_{y,\mathrm{e}}$ sur les mesures de $Y$, alors $\sigma_y = \sigma_{y,\mathrm{e}}$, ce qui se retrouve directement à partir des formules précédentes en écrivant $w_i = 1/\sigma_{y,\mathrm{e}}^2$. On peut vérifier sur les cas correspondants de la figure 11 - cas (a), (b) et (c) - que $\sigma_a$ et $\sigma_b$ sont proportionnelles à $\sigma_{y,\mathrm{e}}$.
▶ Si l'on n'a en revanche pas indiqué d'incertitude sur les mesures de $Y$, alors le programme fait l'hypothèse que $\sigma_y$ vaut une incertitude statistique $\sigma_{y,\mathrm{s}}$ obtenue de la manière suivante : Comme chaque mesure $y_i$ se distribue *a priori* autour de la valeur modèle $a + bx_i$ avec la même incertitude-type $\sigma_y$, les écarts $y_i - (a + bx_i)$ (ce qu'on appelle les *résidus*, voir la figure 12) se distribuent autour de la valeur nulle avec une incertitude-type $\sigma_y$. De la répartition des points de mesure autour de la droite d'équation $y = a + bx$, on peut donc remonter à l'incertitude-type $\sigma_y$ sur les mesures de $Y$. On peut montrer que la meilleure estimation de cette incertitude-type est :

$$\sigma_y = \sigma_{y,\mathrm{s}} = \sqrt{\frac{1}{n-2}\sum_{i=1}^{n}\left[y_i - (a + bx_i)\right]^2}$$

le facteur $n - 2$ venant du fait que l'on a deux paramètres, $a$ et $b$.

### 5.1.3 Accord de données expérimentales avec une loi linéaire

À partir de l'exemple présenté ci-dessus, on discute ici qualitativement de l'accord entre des données expérimentales (incluant éventuellement leurs incertitudes) et la loi linéaire du type $Y = a + bX$ déterminée dans les paragraphes précédents.
▶ Sur la figure 11 (b), les incertitudes sont comparables aux écarts à la droite. Les données et leurs incertitudes sont en bon accord avec une loi linéaire.
▶ Sur la figure 11 (a), les incertitudes sont en moyenne petites par rapport aux écarts à la droite. Soit la loi n'est pas linéaire, soit on a sous-estimé les incertitudes.
▶ Sur la figure 11 (c), les incertitudes sont grandes par rapport aux écarts à la droite. Les données et leurs incertitudes peuvent être modélisées par une loi linéaire. Il est cependant probable qu'on a surestimé les incertitudes.

La discussion qualitative ci-dessus suffit bien souvent, même si le paragraphe 5.1.5 en propose une version quantitative. Il est cependant souvent judicieux de considèrer les résidus $y_i - (a + bx_i)$, tels que représentés sur la figure 12. Si la modélisation des points de mesure par une loi linéaire est justifiée, les résidus se répartissent aléatoirement autour de 0 et sont entièrement décorrelés les uns des autres. Inversement, si l'on essaie de faire une régression linéaire sur des points de mesure qui ne sont en réalité pas issus d'une loi linéaire, comme par exemple sur les panneaux (c) et (d) de la figure 12, les résidus montrent une corrélation d'une valeur à la suivante. Si l'on observe ce genre de comportement, il convient de s'interroger sur la validité du modèle, notamment sur la présence possible d'erreurs systématiques.

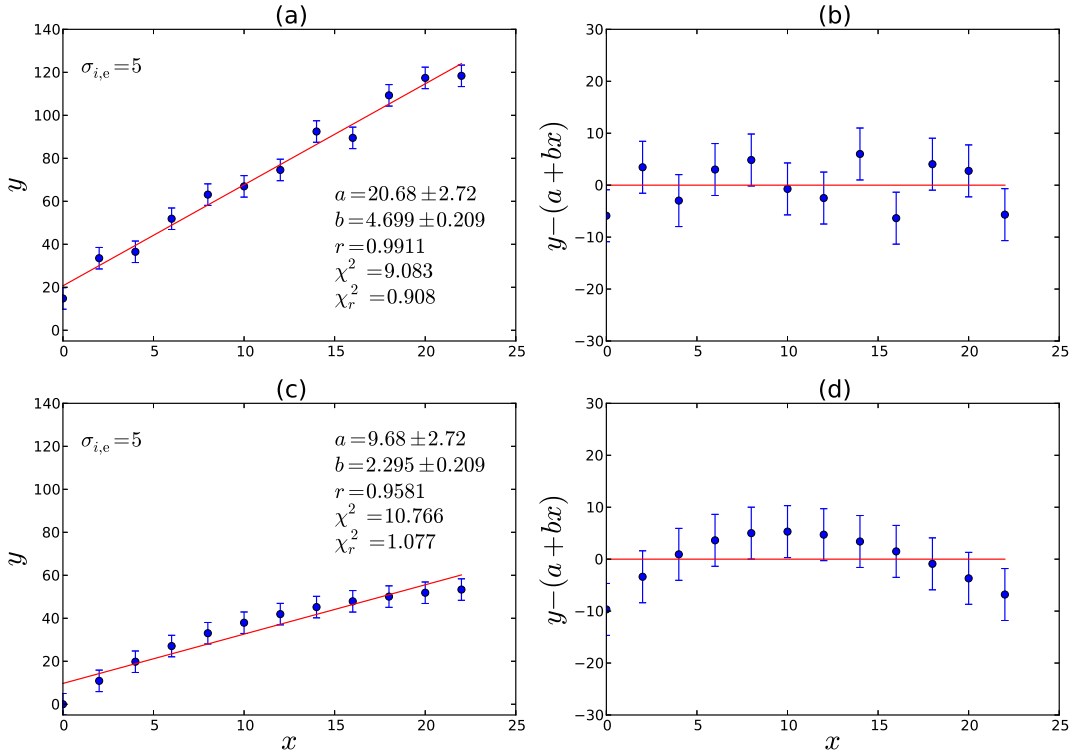


FIGURE 12 – Régressions linéaires $Y = a + bX$ sur l'ensemble des points de mesure du tableau 6 (a), et en ayant remplacé les points de mesure par $y_i = 60\,[1 - \exp(-x_i/10)]$ (c). Les résidus $y_i - (a + bx_i)$ correspondants sont indiqués sur les panneaux (b) et (d).

### 5.1.4 Coefficient de corrélation linéaire

Les programmes d'ajustement donnent généralement la valeur du coefficient de corrélation linéaire [26] :

$$r = \frac{\sum_{i=1}^{n} (x_i - \overline{x})\,(y_i - \overline{y})}{\sqrt{\sum_{i=1}^{n} (x_i - \overline{x})^2}\sqrt{\sum_{i=1}^{n} (y_i - \overline{y})^2}}$$

où $\overline{x}$ et $\overline{y}$ sont respectivement les moyennes empiriques des $x_i$ et des $y_i$. On a $r = 0$ pour un grand nombre de points répartis au hasard, et $r = \pm 1$ pour des points parfaitement alignés [27]. Le coefficient de corrélation $r$ est un outil qui reflète la tendance à la linéarité. En pratique, en TP, quand on fait une régression linéaire, on trouve toujours $r$ très voisin de 1 ou $-1$, comme c'est le cas sur les figures 11 et 12. Ce n'est pas ici un outil efficace pour discuter de la validité d'une loi linéaire. Sous `Qtiplot`, le paramètre $R^2$ donne le carré de ce coefficient. Sous `Igor Pro`, $r$ est donné par la variable $V_{Pr}$, et sous `Synchronie`, c'est `r`.

### 5.1.5 Le $\chi^2$ et le $\chi^2$ réduit

On définit le paramètre $\chi^2$ par [28] :

$$\chi^2 = \sum_{i=1}^{n} \frac{[y_i - (a + bx_i)]^2}{\sigma_i^2}$$

Autrement dit, c'est exactement la quantité $g(a, b)$ introduite plus haut. Dans le cas où tous les $\sigma_i$ sont égaux, c'est la somme des carrés des écarts à la droite divisée par le carré de l'incertitude-type sur $Y$. De manière générale, les paramètres $a$ et $b$ donnés par la méthode des moindres carrés sont ceux qui minimisent le $\chi^2$. Sur la figure 11 - cas (a), (b) et (c) - on voit que le $\chi^2$ est inversement proportionnel au carré de l'incertitude $\sigma_{y,\mathrm{e}}$.

On définit ensuite le $\chi^2$ réduit [29] :

$$\chi_r^2 = \frac{\chi^2}{n-2}$$

Dans cette expression, $n - 2$ est le nombre de degrés de liberté du problème, égal au nombre de points mesurés moins le nombre de paramètres déterminés (ici $a$ et $b$). Notons que si l'on cherche à ajuster une loi $Y = bX$, alors $\chi_r^2 = \chi^2/(n-1)$. Dans le cas où toutes les incertitudes-types sont égales, il est facile d'établir que :

$$\chi_r^2 = \left(\frac{\sigma_{y,\mathrm{s}}}{\sigma_{y,\mathrm{e}}}\right)^2$$

où $\sigma_{y,\mathrm{s}}$ est l'incertitude statistique évaluée par le logiciel (voir le 5.1.2).

Le $\chi^2$ réduit fournit donc un bon critère quantitatif pour décider si des données et leurs incertitudes s'accordent avec une loi linéaire :

▶ Les données expérimentales sont en bon accord avec une loi linéaire si $\chi_r^2 \sim 1$. C'est le cas de la figure 11 (b).

26. Cette formule est valable lorsque tous les $y_i$ ont la même incertitude.
27. Le signe est positif si les $y_i$ augmentent avec les $x_i$, négatif sinon.
28. Sous `Igor Pro`, il est donné par la variable $V_{\mathrm{chisq}}$.
29. Sous `Qtiplot`, cette grandeur est donnée par `Chi2/doF`.

▶ Dans le cas de la figure 11 (a), $\chi^2_r \gg 1$ donc $\sigma_{y,\mathrm{s}} \gg \sigma_{y,\mathrm{e}}$. La loi n'est pas validée ou bien $\sigma_{y,\mathrm{e}}$ a été sous-estimé.
▶ Le cas $\chi^2_r \ll 1$ correspond à la figure 11 (c). Les données peuvent être modélisées par une loi linéaire. Il est cependant probable qu'on a surestimé les incertitudes $\sigma_{y,\mathrm{e}}$.

*Note :* La figure 12 (c) montre cependant que la seule prise en compte du $\chi^2$ réduit est insuffisante pour valider une loi : la répartition des points autour de la droite n'est visiblement pas aléatoire, et pourtant $\chi^2_r = 1,07$ sensiblement égal à celui obtenu sur la figure 12 (a).

*Note :* Certains logiciels renvoient quand même une valeur de $\chi^2$ même si on ne leur a pas fourni les incertitudes $\sigma_{y,\mathrm{e}}$. Cette valeur est calculée en prenant $\sigma_{y,\mathrm{e}} = 1$. On ne peut alors pas tirer grand chose de $\chi^2_r$ si ce n'est la valeur de $\sigma_{y,\mathrm{s}}$. On a alors $\chi^2_r = \sigma^2_{y,\mathrm{s}}$.

## 5.2 Cas général d'un ajustement par une loi $Y = f_\alpha(X)$

Les principes développés dans le cas de la régression linéaire s'appliquent encore dans le cas d'un ajustement par une fonction quelconque $Y = f_\alpha(X)$, dépendant d'un certain nombre de paramètres[30] $\alpha$. La différence essentielle étant qu'il n'y a pas en général de formule donnant directement une estimation optimale des paramètres $\alpha$ à partir des données, contrairement au cas linéaire. Les logiciels d'ajustement déterminent numériquement les valeurs des paramètres qui minimisent le $\chi^2$ défini ici par

$$\chi^2(\alpha) = \sum_{i=1}^{n} \frac{[y_i - f_\alpha(x_i)]^2}{\sigma_i^2}$$

où $n$ est le nombre de points de mesure. On définit également le $\chi^2$ réduit

$$\chi^2_r(\alpha) = \frac{1}{n-p}\chi^2(\alpha)$$

où $p$ est le nombre de paramètres représentés collectivement par $\alpha$.

Contrairement au cas linéaire, il n'y a pas en général de solution analytique à ce problème de minimisation. Il faut donc recourir à des méthodes numériques, et pour cela donner des valeurs initiales aux paramètres $\alpha$. Ceci doit être fait avec une précision suffisante pour que la procédure de minimisation converge. La raison en est que les programmes d'ajustement cherchent un minimum du $\chi^2$ par des méthodes de type "plus fort gradient" en partant des valeurs initiales spécifiées par l'utilisateur. Dans le cas général, il n'est pas garanti que l'hypersurface $z = \chi^2(\alpha)$ ne présente pas de minima locaux. Si de tels minima existent et que l'utilisateur a mal choisi les valeurs initiales des paramètres, la procédure de minimisation aboutira au "mauvais endroit". Même si ces minima n'existent pas, si l'utilisateur a choisi des paramètres trop éloignés des valeurs attendues, il peut être dans un domaine où les gradients du $\chi^2$ sont faibles, de sorte que la procédure ne peut converger avant d'atteindre le nombre maximum d'itérations autorisé.

Par conséquent, en cas d'échec de la procédure :
▶ Contrôler que l'ajustement est fait sur les bonnes variables en ordonnées et abscisses.
▶ Penser à tenir compte d'éventuels décalages sur $X$ et $Y$. Par exemple, au lieu de la fonction $Y = aX^b$, il peut être judicieux d'utiliser $Y = a(X-c)^b + d$. Si cela modifie la forme de la loi qu'on cherche à vérifier (par exemple si $Y = a + bX$ ajuste mieux les données que $Y = bX$ alors que la loi à vérifier est de cette seconde forme), il faut absolument le discuter, notamment du point de vue des possibles erreurs systématiques. Un exemple que vous rencontrerez en TP

30. Dans le cas de la régression linéaire $\alpha = \{a, b\}$.

est celui de l'ajustement de la loi de diffusion du glycérol [31] dans l'eau. Dans cette expérience, représentée sur la Fig. 13, on place une couche de glycérol sous une couche d'eau pure, ce qui provoque l'apparition d'un gradient d'indice optique à l'interface. Ce gradient s'atténue au cours du temps, et on l'observe en faisant passer au travers de la cuve un faisceau laser étalé en une nappe. À l'interface, le gradient d'indice provoque une déviation du faisceau qu'on observe à l'écran. La hauteur $h$ de la déviation est donnée en fonction du temps par

$$h = \frac{(n_g - n_e)\, d}{2\sqrt{\pi D t}}, \tag{1}$$

où $n_g$ et $n_e$ sont les indices optiques du glycérol et de l'eau, $d$ la largeur de la cuve, $D$ le coefficient de diffusion et $t$ le temps. L'ajustement direct de cette formule est voué à l'échec, car le temps $t = 0$ correspondrait à une situation hypothétique dans laquelle les deux fluides seraient parfaitement séparés l'un de l'autre, ce qui n'est pas réalisable en pratique. Il faudra dont remplacer $t$ par $t + t_0$ dans la formule à ajuster, en justifiant l'origine de ce paramètre libre.
▶ Améliorer les valeurs initiales des paramètres. Pour cela, tracer la fonction avec les paramètres initiaux sans opération d'ajustement jusqu'à obtenir un accord visuel suffisant. Ceci peut être indispensable si l'hypersurface $\chi^2(\alpha)$ présente des minima locaux dans l'espace des paramètres.
▶ On peut essayer de bloquer provisoirement la valeur de certains paramètres, de procéder à l'ajustement, puis de libérer un à un les paramètres bloqués.
▶ Réduire la plage des données qu'on cherche à ajuster en effectuant une sélection sur la courbe.

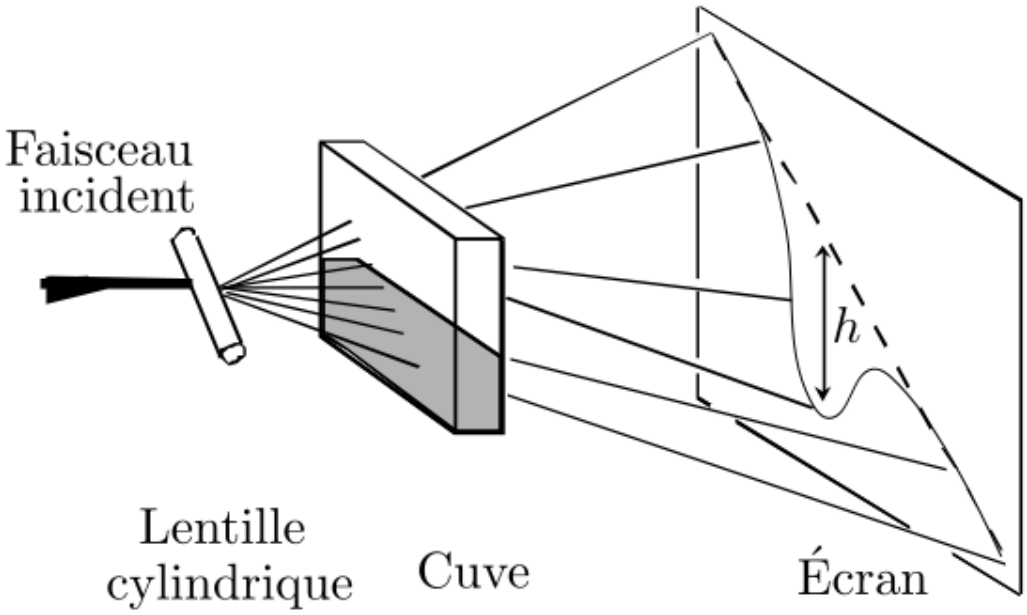


FIGURE 13 – Dispositif expérimental pour la mesure de la diffusion du glycérol dans l'eau.

À titre d'exemple, la figure 14 présente la relation de dispersion $\omega(k)$ mesurée pour des ondes de gravité dans une cuve d'eau de profondeur $h$ et son ajustement par la fonction non-linéaire attendue $f(k) = \sqrt{gk \tanh(kh)}$ avec $g$ l'intensité du champ de pesanteur. Visuellement,

31. On utilise en réalité un mélange eau-glycérol.

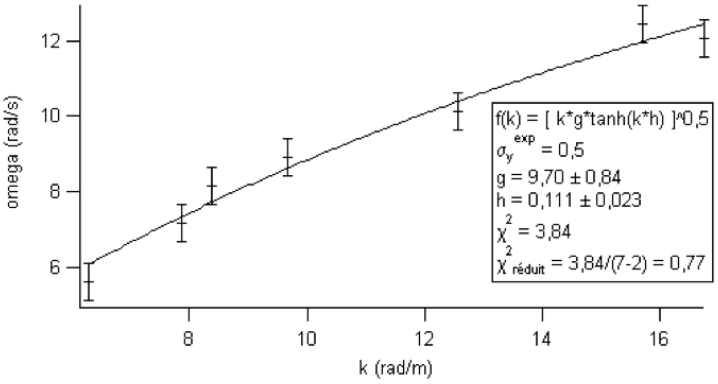


FIGURE 14 – Relation de dispersion des ondes de gravité

on constate que les données et les incertitudes qui leur sont associées sont en accord avec la loi proposée. On obtient une détermination des paramètres $g$ et $h$ avec leurs incertitudes. Par ailleurs, le logiciel évalue le $\chi^2$. On peut en déduire le $\chi^2$ réduit

$$\chi_r^2 = \frac{\chi^2}{5} \simeq 0,77$$

(on a 7 points de mesures et 2 paramètres à déterminer donc 5 degrés de liberté), qui confirme quantitativement l'analyse visuelle.

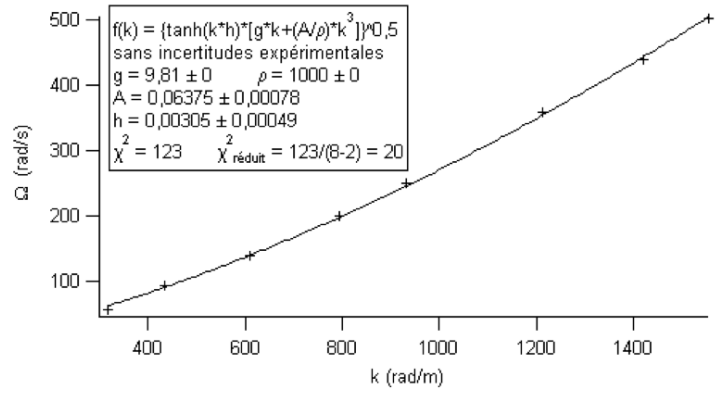


FIGURE 15 – Relation de dispersion des ondes à la surface de l'eau

De manière similaire, la figure 15 présente la relation de dispersion des ondes à la surface de l'eau dans le régime intermédiaire entre le régime capillaire et le régime de gravité. On a réalisé un ajustement par la fonction

$$f(k) = \sqrt{\left(gk + \frac{Ak^3}{\rho}\right) \tanh(kh)}$$

où $A$ et $\rho$ sont respectivement la tension superficielle et la masse volumique de l'eau. On a choisi de bloquer les paramètres $g = 9,81 \pm 0\,\mathrm{m\,s^{-2}}$ et $\rho = 1000 \pm 0\,\mathrm{kg\,m^{-3}}$ pour 2 raisons différentes :

▶ dans le régime étudié, la fonction $f(k)$ est très peu sensible au paramètre $g$ qui serait, si on le laissait libre, déterminé avec une grande incertitude.
▶ la fonction $f$ ne dépend en fait que de 3 paramètres $g$, $h$ et $A/\rho$. Pour déterminer $A$, on a supposé $\rho$ connu.

Ici, on n'a pas rentré d'incertitude-type expérimentale $\sigma_{y,\mathrm{e}}$. On ne peut pas savoir si la loi est validée. Cependant, si on admet cette loi, on a une mesure des paramètres $h$ et $A$ avec leurs incertitudes-types déduites de la répartition statistique des points autour de la courbe (voir la section 5.1.2).

# Références


[1] *Incertitudes et analyse des erreurs dans les mesures physiques*, J.R. Taylor, Editions Dunod, 2000.

[2] *Bruits et Signaux*, D. Pelat, `hal-obspm.ccsd.cnrs.fr/docs/00/09/29/37/PDF/pelat.pdf`

[3] D. W. Hogg, J. Bovy, D. Lang, `https://arxiv.org/abs/1008.4686`

[4] *Probability, Random Variables, and Stochastic Processes*, A. Papoulis, Editions McGraw-Hill, 1984.

[5] *Incertitudes expérimentales*, F.-X. Bally & J.-M. Berroir, BUP, 928, 995, 2010

[6] *Mesure et incertitudes au lycée*, Eduscol, `https://eduscol.education.fr/document/7067/download`, 2021