

# Incertitudes expérimentales

*F.-X. Bally et J.-M. Berroir*

Février 2013

# Table des matières

<b>Introduction</b>	<b>4</b>
<b>1 Erreur et incertitude</b>	<b>4</b>
1.1 Erreurs . . . . .	4
1.1.1 Définition de l'erreur . . . . .	4
1.1.2 Erreurs aléatoires et erreurs systématiques . . . . .	4
1.2 Incertitude . . . . .	5
1.3 Présentation d'un résultat expérimental . . . . .	6
1.4 Comparaison entre valeur mesurée et valeur acceptée . . . . .	7
<b>2 Evaluation des incertitudes par des méthodes statistiques (évaluation de type A)</b>	<b>7</b>
2.1 Analyse statistique d'une série de mesures . . . . .	8
2.1.1 Meilleure estimation de la moyenne de la distribution des valeurs de $x$ . . . . .	8
2.1.2 Meilleure estimation de l'écart-type de la distribution des valeurs de $x$ . . . . .	8
2.1.3 Ecart-type de la moyenne . . . . .	9
2.1.4 Exemple pratique d'analyse statistique d'une série de mesures . . . . .	10
2.2 Propagation des incertitudes . . . . .	10
2.2.1 Cas général : $\mathbf{q} = \mathbf{f}(\mathbf{x}, \mathbf{y}, \dots)$ . . . . .	10
2.2.2 Cas de la somme : $\mathbf{q} = \mathbf{x} + \mathbf{y}$ . . . . .	11
2.2.3 Exemple : $\mathbf{q} = 2/\mathbf{x} + 3/\mathbf{y}^2$ . . . . .	11
2.2.4 Calcul direct de l'incertitude relative . . . . .	11
2.2.5 Exemples pratiques . . . . .	12
2.2.6 Cas où les incertitudes ne sont pas au même niveau de confiance . . . . .	12
2.2.7 Cas d'erreurs aléatoires liées . . . . .	13
<b>3 Vérification d'une loi physique.</b>	
<b>Ajustement de données expérimentales par une fonction</b>	<b>14</b>
3.1 Régression linéaire . . . . .	14
3.1.1 Meilleure estimation des paramètres $a$ et $b$ . . . . .	15
3.1.2 Incertitudes-types $\sigma_a$ et $\sigma_b$ sur les paramètres $a$ et $b$ . . . . .	16
3.1.3 Accord de données expérimentales avec une loi linéaire . . . . .	17

3.1.4	Coefficient de corrélation linéaire . . . . .	18
3.1.5	Le $\chi^2$ . . . . .	18
3.2	Cas général d'un ajustement par une loi $y = f(x)$ . . . . .	19

**Annexe : Rappels élémentaires de statistique** **21**

A.1	Distribution de probabilité d'une grandeur aléatoire . . . . .	21
A.2	Valeur moyenne et écart-type . . . . .	21
A.3	Exemples de distributions de probabilité . . . . .	21
A.4	Importance physique de la Loi Normale . . . . .	23

# Introduction

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. Cet article vise à fournir les outils nécessaires à l'analyse de résultats expérimentaux. On y introduit d'abord la notion d'incertitude. Puis, on montre comment, à l'aide d'une étude statistique, on peut quantifier les incertitudes associées au caractère aléatoire des processus de mesure. On montre ensuite comment évaluer l'incertitude sur une grandeur obtenue par calcul à partir d'un certain nombre d'autres grandeurs mesurées. Enfin, on aborde le problème de la modélisation et de la vérification d'une loi physique. On fournit les outils nécessaires à une approche quantitative de ce problème. Dans tous les cas, l'accent est mis sur la pratique et de nombreux exemples concrets sont présentés.

## 1 Erreur et incertitude

Beaucoup de scientifiques confondent ces deux termes et parlent de calculs d'erreurs au lieu de calculs d'incertitudes.

### 1.1 Erreurs

#### 1.1.1 Définition de l'erreur

Lors de la mesure d'une grandeur physique  $x$ , l'erreur est la différence entre la valeur mesurée<sup>1</sup>  $x$  et la valeur vraie  $X$ . *La valeur vraie est en général inconnue* (puisqu'on la cherche).

#### 1.1.2 Erreurs aléatoires et erreurs systématiques

##### a. Erreurs aléatoires

Lorsqu'on mesure la période d'oscillation d'un pendule en opérant avec un chronomètre manuel, on constate qu'en répétant les mesures on trouve des résultats légèrement différents, dus surtout aux retards de déclenchement qui vont réduire ou accroître la valeur de la période suivant qu'ils ont lieu au début ou à la fin de la mesure. Ce phénomène sera détecté par une étude statistique. On parle d'erreur aléatoire. Le résultat de la mesure est caractérisé par une distribution de probabilité<sup>2</sup> répartie autour de la valeur vraie dans le cas d'erreurs purement aléatoires.

##### b. Erreurs systématiques

Supposons maintenant qu'on mesure la période d'oscillation d'un pendule avec un chronomètre faussé qui indique toujours des temps 2 % trop faibles. L'étude statistique ne le détectera pas. On parle d'erreur systématique : c'est la composante de l'erreur qui ne varie pas dans des conditions de mesure répétées. Plus généralement les erreurs systématiques ont des origines diverses :

- erreur d'étalonnage

*Exemple* : Millikan a trouvé une valeur inexacte de la charge de l'électron parce qu'il avait pris une valeur fautive de la viscosité de l'air.

- oubli d'un paramètre

---

1. Pour simplifier les notations, dans toute la suite on désigne par la même lettre  $x$  la grandeur physique et sa valeur mesurée.

2. Des rappels élémentaires de statistiques sont présentés en annexe.

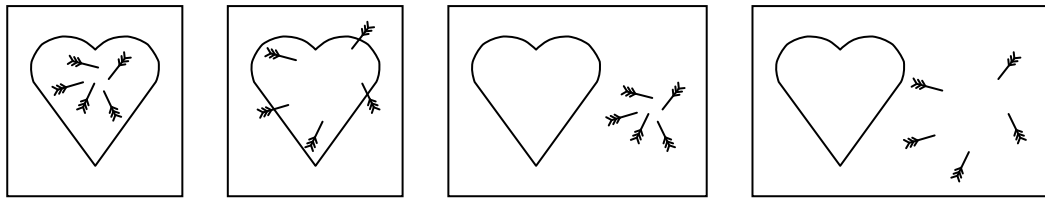
*Exemple* : influence de la température sur la vitesse du son (si on ne précise pas la température il est impossible de comparer la mesure à une valeur de référence)

- procédure erronée

*Exemple* : mesure d'une résistance sans tenir compte des résistances de l'ampèremètre et du voltmètre...

Les erreurs systématiques sont difficile à détecter à 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).

On représente classiquement les rôles respectifs des erreurs aléatoires et systématiques par une analogie avec un tir sur cible (cf. figure 1), le centre de la cible représentant la valeur vraie de la grandeur à mesurer :



**Figure 1** - Rôles respectifs des erreurs aléatoires et systématiques

- si tous les impacts sont proches du centre : faibles erreurs aléatoires et faible erreur systématique

- si les impacts sont très étalés mais centrés en moyenne sur la cible : fortes erreurs aléatoires et faible erreur systématique

- si les impacts sont groupés mais loin du centre : faibles erreurs aléatoires et forte erreur systématique

- si les impacts sont étalés et loin du centre : fortes erreurs aléatoires et forte erreur systématique.

Le défaut de cette analogie est qu'en général, dans les mesures physiques on ne connaît pas le centre de la cible !

*Un exemple plus complexe* : si on mesure une distance avec une règle en métal souple, la flexion de la règle va introduire une erreur systématique (la distance lue est toujours trop grande) et aléatoire (la flexion de la règle est variable).

## 1.2 Incertitude

L'incertitude  $\delta x$  traduit les tentatives scientifiques pour estimer l'importance de l'erreur aléatoire commise. En absence d'erreur systématique, elle définit un intervalle autour de la valeur mesurée qui inclut la valeur vraie avec un niveau de confiance déterminé. La détermination de l'incertitude n'est pas simple à priori. **On rencontre en pratique deux situations :**

-  $\delta x$  est évalué statistiquement : on parle alors d'évaluation de type A de l'incertitude.

On cherche dans ce cas à caractériser la distribution de probabilité des valeurs de  $x$ , en évaluant le mieux possible la valeur moyenne et l'écart-type de cette distribution. Ceci se fait par l'analyse statistique d'un ensemble de mesures de  $x$ , présentée en détail au paragraphe 2. En l'absence d'erreur systématique, l'estimation de la valeur moyenne est la meilleure estimation de la valeur vraie  $X$  tandis que l'incertitude  $\delta x$ , directement reliée à l'estimation de l'écart-type de la distribution, définit un intervalle dans lequel la valeur vraie de  $X$  se trouve avec un niveau de confiance connu. On choisit le plus souvent comme incertitude l'estimation de l'écart-type de la distribution. On parle alors d'*incertitude-type*.

- $\delta x$  est évalué par d'autres moyens : on parle alors d'évaluation de type B de l'incertitude.

Si on ne dispose pas du temps nécessaire pour faire une série de mesures, on estime  $\delta x$  à partir des spécifications des appareils de mesures et des conditions expérimentales.

*Exemple 1 - Utilisation d'un voltmètre numérique* : en ouvrant la notice d'un voltmètre numérique, on trouve typiquement les indications suivantes : la précision  $\Delta$  de la mesure est donnée par  $\pm 2$  fois le dernier digit  $\pm 0,1\%$  de la valeur lue. On peut considérer que l'indication donnée par le fabricant a 2 origines<sup>3</sup> :

- celle qui provient d'une erreur de calibrage (variable d'un appareil à l'autre, d'un calibre à l'autre...). Cette erreur est systématique quand on utilise le même calibre d'un même appareil. Elle devient aléatoire quand on utilise plusieurs calibres ou plusieurs appareils, même de modèles identiques. Elle est essentiellement présente dans les  $\pm 0,1\%$  de la valeur lue.

- celle qui provient d'erreurs aléatoires (bruit...). Elle est essentiellement présente dans les  $\pm 2$  fois le dernier digit.

Pour évaluer l'incertitude-type sur une mesure, on procède à une évaluation de type B. Deux cas se présentent :

- le constructeur donne un niveau de confiance, en précisant par exemple que  $\Delta$  correspond à 3 écart-types. On en déduit facilement l'écart-type  $\delta x$ . Dans ce cas, le fabricant s'engage sur la précision de son appareil<sup>4</sup>.

- 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 à l'écart-type  $\delta x = \Delta/\sqrt{3}$ . C'est le cas de la plupart des appareils de mesure courants.

*Exemple 2* : dans une expérience d'interférences avec les fentes d'Young, on mesure la distance  $d$  (de l'ordre du mètre) entre les bifentes 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<sup>5</sup>. Et si 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.

### 1.3 Présentation d'un résultat expérimental

L'écriture rapportant la mesure d'une grandeur physique  $x$  est :

valeur mesurée de  $x = \bar{x} \pm \delta x$

où  $\bar{x}$  est la meilleure estimation de la valeur vraie  $X$  et  $\delta x$  l'incertitude-type sur la mesure (incertitude absolue).

En l'absence d'erreurs systématiques, la valeur vraie de  $x$  se trouve probablement dans (ou proche de)

3. cf. 2.2.7

4. Norme NFC 42670

5. Pour formaliser ce résultat, on peut supposer, comme pour le voltmètre de l'exemple précédent, que la répétition de la mesure conduirait à une distribution de probabilité uniforme centrée sur la valeur vraie et de demi-largeur  $0,5$  cm. Le calcul de l'écart-type de cette distribution donne une incertitude-type de  $0,5/\sqrt{3} = 0,29$  cm.

l'intervalle allant de  $\bar{x} - \delta x$  à  $\bar{x} + \delta x$ .

On définit aussi l'incertitude-type relative ou fractionnaire :  $\delta x/|\bar{x}|$

*Exemple* : dans une expérience d'ondes stationnaires en acoustique on mesure la distance entre 7 nœuds de pression, on trouve 862 mm avec une incertitude-type absolue estimée à 20 mm (parce qu'on a du mal à repérer précisément les nœuds). A la calculatrice on obtient pour la longueur d'onde et pour l'incertitude-type absolue les résultats suivants :  $862 / 3 = 287,333$  mm et  $20 / 3 = 6,667$  mm.

- Compte tenu du nombre de chiffres significatifs de la mesure initiale on garde :  $\lambda = 287,3$  mm.  
- L'incertitude-type étant toujours évaluée grossièrement, on garde 1 à 2 chiffres significatifs :  $\delta\lambda = 7$  mm.

- Le dernier chiffre significatif de la mesure doit être cohérent avec l'incertitude-type, on écrira donc :  
valeur mesurée de  $\lambda = 287 \pm 7$  mm ou valeur mesurée de  $\lambda = 287 \text{ mm} \pm 2\%$

## 1.4 Comparaison entre valeur mesurée et valeur acceptée

Ayant obtenu la valeur mesurée avec son intervalle d'incertitude, on la compare à la valeur de référence (pour une valeur expérimentale de référence, ne pas parler de valeur exacte, parler de *valeur tabulée*). Il n'est pas anormal que l'intervalle ne contienne pas la valeur de référence. Ainsi dans le cas fréquent d'une distribution Gaussienne, le tableau 4 donné en annexe 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 atteint plus de  $2 \delta x$  (probabilité que la mesure soit bonne : 1 chance sur 20 à  $2 \delta x$  et 1 chance sur 100 à  $2,5 \delta x$  pour une distribution Gaussienne). Si c'est le cas il faut invoquer des erreurs systématiques.

*Exemple pratique* : on mesure la vitesse du son dans l'air à 20°C. On effectue une série de mesures qui conduit à  $\bar{v} = 335 \text{ m.s}^{-1}$  et à une incertitude-type  $\delta v = 5 \text{ m.s}^{-1}$ . On a donc : valeur mesurée de  $v = 335 \pm 5 \text{ m.s}^{-1}$ . La valeur tabulée indique à cette température :  $v = 343 \text{ m.s}^{-1}$ . Elle est en-dehors du domaine d'incertitude, mais l'écart entre 343 et 335 vaut  $1,6 \delta v$ . La table donnée en annexe (paragraphe A.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.

Il n'est pas possible de tenir compte des erreurs systématiques 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 tabulée<sup>6</sup>.

## 2 Evaluation des incertitudes par des méthodes statistiques (évaluation de type A)

L'évaluation d'une incertitude expérimentale à partir de l'étude statistique directe de la grandeur physique recherchée est présentée dans la partie 2.1. Dans la partie 2.2, on donne les méthodes de calcul d'une incertitude dans le cas où une formule mathématique relie la grandeur physique étudiée aux diverses grandeurs effectivement mesurées.

---

6. A contrario, le fait de trouver un bon accord ne prouve pas qu'il n'y a pas d'erreur systématique : il peut y en avoir plusieurs qui se compensent.

## 2.1 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. Dans les deux premiers paragraphes, on décrit les méthodes statistiques qui permettent d'évaluer la valeur vraie de  $x$  et l'incertitude-type. Dans le paragraphe suivant, on montre comment réduire l'incertitude sur la détermination de  $x$ .

Le traitement statistique est basé sur la répétition des mesures de  $x$ . Si on était capable de réaliser une infinité de mesures, on déterminerait la distribution de probabilité de  $x$ , en particulier sa valeur moyenne  $X$  (ou valeur vraie) et son écart type  $\sigma$ . Dans la pratique, on réalise un nombre fini  $n$  de mesures, de résultats respectifs  $x_1, x_2, \dots, x_n$ , dont on cherche à extraire les meilleures estimations de  $X$  et  $\sigma$ . Les méthodes statistiques qui permettent d'obtenir ces meilleures estimations sont présentées dans les paragraphes suivants, sans démonstration.

### 2.1.1 Meilleure estimation de la moyenne de la distribution des valeurs de $x$

La meilleure estimation de la valeur vraie  $X$ , notée  $\bar{x}$ , obtenue à partir des  $n$  mesures  $x_1, x_2, \dots, x_n$  est la moyenne de ces mesures<sup>7</sup>.

$$\text{Meilleure estimation de } X = \bar{x} = \frac{x_1 + x_2 + \dots + x_n}{n}$$

### 2.1.2 Meilleure estimation de l'écart-type de la distribution des valeurs de $x$

La meilleure estimation de  $\sigma$  déduite des  $n$  mesures  $x_1, x_2, \dots, x_n$ , notée  $\sigma_x$ , est donnée par<sup>7</sup> :

$$\text{Meilleure estimation de } \sigma = \sigma_x = \sqrt{\frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})^2}$$

Note : Le facteur  $n - 1$ , et non pas  $n$ , vient du fait que la formule ci-dessus utilise  $\bar{x}$ , seule quantité accessible à l'expérience, et non pas  $X$ . 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.

Il est maintenant légitime de se poser la question de la précision de ces estimations, en particulier de celle de la moyenne.

---

7. La démonstration de ce résultat est assez simple dans le cas où la grandeur à mesurer a une distribution de probabilité gaussienne (cf. *Incertitudes et analyse des erreurs dans les mesures physiques*, paragraphe 5.5, p. 128 et annexe E1, p.280, J.R. Taylor, Editions Dunod, 2000). Les meilleures estimations pour  $X$  et  $\sigma$  maximisent la probabilité  $P_{X,\sigma} \propto 1/\sigma^n e^{-\sum(x_i - X)^2/2\sigma^2}$  d'obtenir les résultats  $x_1, \dots, x_n$  lors de  $n$  mesures. On obtient  $\bar{x}$  (resp.  $\sigma_x$ ) en dérivant  $P_{X,\sigma}$  par rapport à  $X$  (resp.  $\sigma$ ) à  $x_i$  fixés.



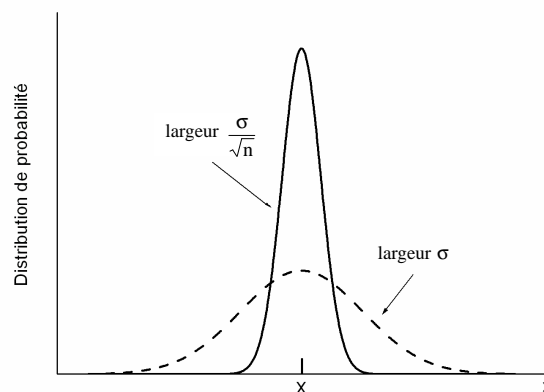
### 2.1.3 Ecart-type de la moyenne

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

$$\text{Incertitude - type sur } \bar{x} = \sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}}$$

$\sigma_{\bar{x}}$  représente l'incertitude-type sur la détermination de la valeur vraie  $X$  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 2). Dans la pratique,  $\sqrt{n}$  croît lentement et améliorer la précision d'un facteur 10 oblige à effectuer 100 fois plus de mesures.

**Figure 2** - Lorsqu'on effectue une mesure unique, la valeur trouvée suit la distribution de probabilité représentée en pointillés.  $n$  mesures indépendantes se répartissent aléatoirement sur cette courbe. Lorsqu'on prend la valeur moyenne de ces  $n$  mesures, les écarts à la valeur vraie se compensent statistiquement, avec d'autant plus d'efficacité que  $n$  est grand. Si on réalise plusieurs déterminations de la moyenne de  $n$  mesures, elles suivent la distribution en trait plein.



Un exemple simple pour illustrer cette propriété : si on a besoin d'une résistance de  $1 \text{ k}\Omega$  il vaut mieux utiliser 10 résistances de  $100 \Omega$  à 1% en série plutôt qu'une seule résistance de  $1000 \Omega$  à 1%.

En pratique  $\sigma$  n'est pas connu et on utilise son évaluation  $\sigma_x$  pour déterminer  $\sigma_{\bar{x}}$ . Dans le cas où on effectue une moyenne sur un petit nombre de mesures,  $\sigma_x$  n'est pas une bonne estimation de  $\sigma$  et la théorie montre qu'il faut alors utiliser  $\sigma_{\bar{x}} = t \times \frac{\sigma_x}{\sqrt{n}}$  où  $t$  est appelé *coefficient de Student*. Ce coefficient, qui est tabulé, dépend du nombre de points moyennés et du niveau de confiance<sup>8</sup>. En TP usuel, 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 l'exemple du paragraphe suivant).

En résumé, si on réalise  $n$  mesures de  $x$ , avec les résultats  $x_1, x_2, \dots, x_n$ , on écrira le résultat final sous la forme :

$$x = \bar{x} \pm \frac{\sigma_x}{\sqrt{n}}$$

où  $\bar{x}$  et  $\frac{\sigma_x}{\sqrt{n}}$  sont les meilleures estimations de la valeur vraie et de l'incertitude-type.

8. Ici on utilise un niveau de confiance correspondant à un écart-type (68%), certains auteurs préfèrent utiliser un niveau de confiance correspondant à 2 écart-types (95%)

### 2.1.4 Exemple pratique d'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 :

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

**Table 1** - Mesures indépendantes de la raie verte du mercure

En utilisant un logiciel de traitement des données<sup>9</sup>, on obtient :  $\bar{\lambda} = 548,04$  nm et  $\sigma_{\lambda} = 9,72$  nm. On en déduit que l'incertitude sur la moyenne des 8 mesures vaut :  $\delta\lambda = \sigma_{\bar{\lambda}} = 9,72/\sqrt{8} = 3,44$  nm. On peut donc écrire : meilleure estimation de  $\lambda = 548 \pm 3$  nm ou  $545 \leq \lambda \leq 551$  nm<sup>10</sup>. On peut comparer à la valeur tabulée :  $\lambda_{tab} = 545,07$  nm et conclure qu'il y a une bonne concordance.

*Remarque* : L'incertitude-type  $\delta\lambda$  est calculée à partir de  $\sigma_{\lambda}$  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 paragraphe 2.1.3 permet de tenir compte de cet effet. Le tableau 2 donne les valeurs du coefficient  $t$  en fonction du nombre de mesures pour un intervalle de confiance de 68%.

**Table 2** - Coefficient de Student

nombres de mesures $n$	2	3	4	5	6	7	8	9	10	20	40	$\infty$
Coefficient de Student $t$	1,84	1,32	1,20	1,14	1,11	1,09	1,08	1,07	1,06	1,03	1,01	1,00

Le tableau 2 donne  $t = 1,08$  pour 8 mesures (pour un nombre infini de mesures,  $t = 1$ ).

On doit donc écrire en toute rigueur :  $\lambda = \bar{\lambda} \pm t \times \frac{\sigma_{\lambda}}{\sqrt{n}} = 548,04 \pm 1,08 \times 9,72/\sqrt{8} = 548 \pm 4$  nm. Cet exemple montre qu'on peut oublier la correction de Student dans le cadre des TP usuels.

## 2.2 Propagation des incertitudes

On s'intéresse ici au problème suivant : on connaît les grandeurs expérimentales  $x, y, \dots$  avec les incertitudes-types  $\delta_x, \delta_y, \dots$ <sup>11</sup> Quelle est l'incertitude-type  $\delta_q$  sur la grandeur  $q = f(x, y, \dots)$  ?

### 2.2.1 Cas général : $q = f(x, y, \dots)$

Si les erreurs sur  $x$  et  $y$  sont indépendantes, on peut montrer que<sup>12</sup> :

$$\delta_q = \sqrt{\left[\frac{\partial f}{\partial x}\right]^2 (\delta_x)^2 + \left[\frac{\partial f}{\partial y}\right]^2 (\delta_y)^2 + \dots}$$

9. - Sous Igor, utiliser (Analysis  $\rightarrow$  Statistics :  $V_{avg} = \bar{x}$  et  $V_{sdev} = \sigma_x$ )

- Sous Synchronie, le module *Statistiques* ne donne pas  $\sigma_x$  mais  $\sqrt{\frac{n-1}{n}}\sigma_x$ , qui est en pratique très voisin de  $\sigma_x$

- Sous Excel, utiliser (Insertion  $\rightarrow$  Fonction : MOYENNE =  $\bar{x}$  et ECARTYPE =  $\sigma_x$ )

- Sous Kaleidagraph, utiliser (Fonctions  $\rightarrow$  Statistics : Mean et Std Deviation)

10. Remarquer qu'on arrondit 3,44 par 3 qui est sa plus proche valeur avec un seul chiffre significatif. Dans les anciens calculs d'incertitude, où l'incertitude était un majorant de l'erreur, on aurait arrondi à la valeur supérieure.

11. Par exemple, si on a mesuré  $n$  valeurs de  $x$ , on a  $\delta_x = \sigma_{\bar{x}}$ .

12. On trouvera une justification de ce résultat dans *Incertitudes et analyse des erreurs dans les mesures physiques*, Paragraphe 5.6, p. 132, J.R. Taylor, Editions Dunod, 2000.

On dit que les incertitudes-types s'ajoutent quadratiquement <sup>13</sup>.

### 2.2.2 Cas de la somme : $q = x + y$

$$\delta_q = \sqrt{(\delta_x)^2 + (\delta_y)^2}$$

L'addition quadratique donne un résultat plus faible que l'addition en module :  $\sqrt{(\delta_x)^2 + (\delta_y)^2} \leq \delta_x + \delta_y$ . On peut le comprendre physiquement : la probabilité pour que  $x + y$  s'écarte de sa valeur moyenne de plus d'un écart type est plus faible que la somme des probabilités correspondantes pour  $x$  et  $y$ . En effet,  $x$  et  $y$  peuvent s'écarter simultanément de leur valeur moyenne de plus d'un écart-type sans qu'il en soit de même pour leur somme.

*Application* : le résultat ci-dessus permet de retrouver la formule, donnée au paragraphe 2.1.3, reliant l'écart-type de la valeur moyenne de  $n$  mesures indépendantes de la même variable à l'écart-type de la variable :  $\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}}$

### 2.2.3 Exemple : $q = 2/x + 3/y^2$

Pour calculer  $\delta q$ , on prend la différentielle des 2 membres :

$$dq = (-2/x^2) dx + (-6/y^3) dy.$$

On remplace  $d$  par  $\delta$ , l'addition par une addition quadratique et on prend la racine carrée pour obtenir :

$$\delta_q = \sqrt{[-2/x^2]^2 (\delta_x)^2 + [-6/y^3]^2 (\delta_y)^2}$$

### 2.2.4 Calcul direct de l'incertitude relative

Lorsque  $q$  contient des produits, rapports ou puissances, on a intérêt à utiliser la différentielle logarithmique pour exprimer directement l'incertitude relative sur  $q$  en fonction des incertitudes sur  $x, y, \dots$

*Premier Exemple* :  $q = -3xy^2/z$

On prend le logarithme népérien de la valeur absolue des deux membres :

$$\ln|q| = \ln|-3| + \ln|x| + 2\ln|y| - \ln|z|$$

puis on calcule la différentielle :  $dq/q = (1/x) dx + (2/y) dy + (-1/z) dz$

On en déduit :  $\delta_q/|q| = \sqrt{[1/x]^2 (\delta_x)^2 + [2/y]^2 (\delta_y)^2 + [-1/z]^2 (\delta_z)^2}$

*Deuxième exemple* :  $q = (x - y)/(x + y)$  (on suppose  $x - y > 0$  et  $x + y > 0$ )

Cet exemple est intéressant car les influences de  $x$  au numérateur et au dénominateur se compensent partiellement (si  $x$  augmente, numérateur et dénominateur augmentent donc  $q$  varie peu). La différentielle logarithmique conduit à :

---

13. De même qu'en optique, on additionne les intensités de 2 sources incohérentes et pas leurs amplitudes.

$$dq/q = [1/(x-y)] dx - [1/(x-y)] dy - [1/(x+y)] dx - [1/(x+y)] dy.$$

La formule de propagation des incertitudes n'étant valable que pour des variables indépendantes, il faut impérativement mettre en facteur  $dx$  et  $dy$  avant d'évaluer  $\delta q$  :

$$dq/q = [1/(x-y) - 1/(x+y)] dx + [-1/(x-y) - 1/(x+y)] dy$$

d'où :

$$\delta_q/|q| = \sqrt{[1/(x-y) - 1/(x+y)]^2 (\delta_x)^2 + [-1/(x-y) - 1/(x+y)]^2 (\delta_y)^2}$$

La compensation évoquée ci-dessus se traduit par le signe - dans le terme en facteur de  $(\delta_x)^2$ .

### 2.2.5 Exemples pratiques

- a. On détermine la période  $T$  d'un pendule en chronométrant manuellement 20 oscillations. On déclenche et arrête le chronomètre aux instants  $t_1$  et  $t_2$  où le pendule passe par la position d'équilibre. On a  $T = (t_2 - t_1)/20$ . On a mesuré  $t_1 = 0$  et  $t_2 = 13,8$  s avec  $\delta_{t_1} = \delta_{t_2} = 0,3$  s. La formule générale donne :  $\delta_T = \sqrt{[1/20]^2 (\delta_{t_2})^2 + [-1/20]^2 (\delta_{t_1})^2} = 0,02$  s. Si on admet que la distribution de probabilité est gaussienne, on peut dire qu'il y a 2 chances sur 3 (68%) pour que l'erreur aléatoire sur  $T$  soit inférieure à cette valeur (cf. table 4, page 22).
- b. 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 :  $\delta_P/P = \sqrt{(\delta_R/R)^2 + 4(\delta_I/I)^2} = \sqrt{4 \cdot 10^{-3} + 2 \cdot 10^{-4}} \approx \sqrt{4 \cdot 10^{-3}} \approx 0,06$  ou 6%. On remarque que l'incertitude sur l'intensité est négligeable. **De manière générale, il est important de comparer les incertitudes des différentes grandeurs mesurées.** Il est clair que dans l'exemple ci-dessus, la principale cause d'erreur porte sur la résistance et que si 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.
- c. Pour trouver la capacité thermique massique  $c$  d'un liquide lors d'une expérience de calorimétrie, on mesure l'élévation de température d'une certaine quantité de ce liquide dans laquelle on dissipe une énergie électrique connue. On obtient la capacité thermique cherchée par la formule :  $c = UIt/[m(\theta_2 - \theta_1)]$  où  $U$ ,  $I$ ,  $t$ ,  $m$  et  $(\theta_2 - \theta_1)$  représentent respectivement une tension, une intensité, une durée, une masse et une variation de température. On a estimé :  $\delta_U/U = 1\%$ ,  $\delta_I/I = 2\%$ ,  $t = 153$  s,  $\delta_T = 2$  s,  $m = 345$  g,  $\delta_m = 5$  g,  $\theta_1 = 19,3^\circ C$ ,  $\theta_2 = 20,3^\circ C$  et  $\delta_\theta = 0,1^\circ C$ . Le calcul d'incertitude montre que la seule incertitude qui joue un rôle est celle sur les températures ; on obtient finalement :  $\delta_c/c \approx \sqrt{2} \delta_\theta/(\theta_2 - \theta_1) = 14\%$ . Les commentaires du paragraphe précédent s'appliquent encore ici. On constate aussi qu'il faut si possible éviter l'apparition de petites différences : l'incertitude relative sur la différence de 2 nombres très voisins est grande. Pour améliorer cette expérience il faudrait accroître nettement la variation de température en augmentant la durée de l'expérience et/ou l'intensité du courant.

### 2.2.6 Cas où les incertitudes ne sont pas au même niveau de confiance

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%<sup>14</sup>. D'autres utilisent *l'incertitude-élargie* qui correspond à 2 écarts-types soit un niveau de confiance d'environ 95%<sup>15</sup>.

14. en supposant la statistique gaussienne.

15. En réalité, la norme définissant les incertitudes des instruments est plus complexe que ce qui est présenté ici (cf. BUP 889, cahier n° 2, p. 219, *Expression et évaluation des incertitudes de mesures*, J.-L. Bretonnet).

*Exemple* : on détermine une puissance par la relation  $P = UI$ . On mesure  $I = 1,27 \text{ A}$  avec une incertitude-type égale à  $0,02 \text{ A}$ . On mesure  $U = 15,24 \text{ V}$  avec une incertitude-élargie égale à  $0,06 \text{ V}$ . On peut donc poser :  $\delta_I = 0,02 \text{ A}$  et comme l'incertitude-élargie vaut le double de l'incertitude-type :  $\delta_U = 0,03 \text{ V}$ .

On en déduit l'incertitude-type sur  $P$  :  $\delta_P/P = \sqrt{(0,02/1,27)^2 + (0,03/15,24)^2} \approx 2\%$ .

### 2.2.7 Cas d'erreurs aléatoires liées

La loi de propagation des incertitudes indépendantes suppose que les variables  $x, y, \dots$  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 :

$$\delta q \leq \left| \frac{\partial q}{\partial x} \right| \delta x + \left| \frac{\partial q}{\partial y} \right| \delta y + \dots$$

On va voir cependant sur un exemple pratique que cette inégalité risque d'être de peu d'intérêt.

*Exemple pratique* : 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). Il faut comprendre que pour l'utilisateur d'un appareil, il s'agit d'une incertitude aléatoire (sauf s'il procédait à un étalonnage de son appareil, ce qu'il ne fait pas, ne disposant pas d'un appareil de référence).

L'utilisateur mesure 2 tensions voisines :  $V_1 = 12,71 \text{ V}$  et  $V_2 = 9,32 \text{ V}$  afin de déterminer  $V_3 = V_1 - V_2$  et  $V_4 = V_1 + V_2$ .

- S'il utilise 2 voltmètres, l'un pour mesurer  $V_1$ , l'autre pour mesurer  $V_2$ , les incertitudes aléatoires  $\delta V_1 = 12,71 \times 0,01 = 0,13 \text{ V}$  et  $\delta V_2 = 9,32 \times 0,01 = 0,09 \text{ V}$  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 :  $dV_3 = dV_1 - dV_2$  avec  $dV_2/V_2 = dV_1/V_1$  puisqu'il y a uniquement une erreur de calibrage. On en déduit  $dV_3 = dV_1(1 - V_2/V_1)$  puis, en passant aux incertitudes,  $\delta V_3 = \delta V_1 |1 - V_2/V_1| = 0,035 \text{ V}$  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 retranchent. De même, on trouve  $\delta V_4 = \delta V_1 |1 + V_2/V_1| = 0,22 \text{ V}$ . Pour  $V_4$ , l'incertitude obtenue avec un voltmètre est supérieure à celle obtenue avec 2 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.

### 3 Vérification d'une loi physique.

#### Ajustement de données expérimentales par une fonction

On cherche souvent à vérifier une loi physique reliant 2 grandeurs distinctes. Ainsi, dans une expérience sur les semiconducteurs, on est amené à vérifier que la tension de Hall  $V_H$  aux bornes d'un échantillon semiconducteur est reliée de manière linéaire au champ magnétique  $B$  par  $V_H = (I/qbn_p)B$  où  $I$  est le courant qui traverse l'échantillon,  $q$  la charge d'un porteur,  $b$  l'épaisseur dans la direction de  $B$  et  $n_p$  est le nombre de porteurs par unité de volume du semiconducteur. En mesurant  $V_H$  pour différentes valeurs de  $B$  à  $I$  fixé, 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 2 grandeurs  $x$  et  $y$  sont reliées par une loi du type  $y = f(x)$  où  $f$  est une fonction dépendant d'un certain nombre de paramètres. On mesure une série de  $n$  valeurs  $x_1, x_2, \dots, x_n$  et les valeurs correspondantes  $y_1, y_2, \dots, y_n$ . L'objet des paragraphes suivants est de montrer comment on détermine les meilleures valeurs des paramètres de la fonction  $f$  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  $f$  ainsi déterminée s'ajuste aux données expérimentales.

#### 3.1 Régression linéaire

Dans de nombreux cas, on cherche à savoir dans quelle mesure des données expérimentales s'accordent avec une loi linéaire<sup>16</sup> 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 très 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).

On dispose donc d'un tableau de  $n$  mesures  $(x_1, y_1), (x_2, y_2), \dots, (x_n, y_n)$  et éventuellement pour chacune de ces mesures, de l'incertitude associée à la mesure de  $y$  (on note alors  $\sigma_i^{exp}$  l'incertitude-type sur  $y_i$ ). On commence par représenter graphiquement les différents couples de points. A titre d'exemple, on a représenté sur la figure 3.a les données du tableau 3 (ici  $n = 12$ ).

**Table 3 - Données utilisées pour les exemples de régression linéaire**

<b>x</b>	0	2	4	6	8	10	12	14	16	18	20	22
<b>y</b>	14,79	33,52	36,50	51,88	63,11	66,94	74,58	92,46	89,50	109,29	117,40	118,37

On appelle ensuite le programme de modélisation mathématiques (ou fit<sup>17</sup>) du logiciel utilisé<sup>18</sup>. Celui-ci trace la meilleure droite passant par les points en utilisant les résultats du paragraphe suivant.

16. Le cas d'une loi du type  $y = bx$  n'est pas détaillé ici car la démarche est exactement la même. Les deux types d'ajustement linéaire sont proposés par tous les programmes de traitement de données.

17. En anglais, fit = ajuster, mettre à la bonne dimension.

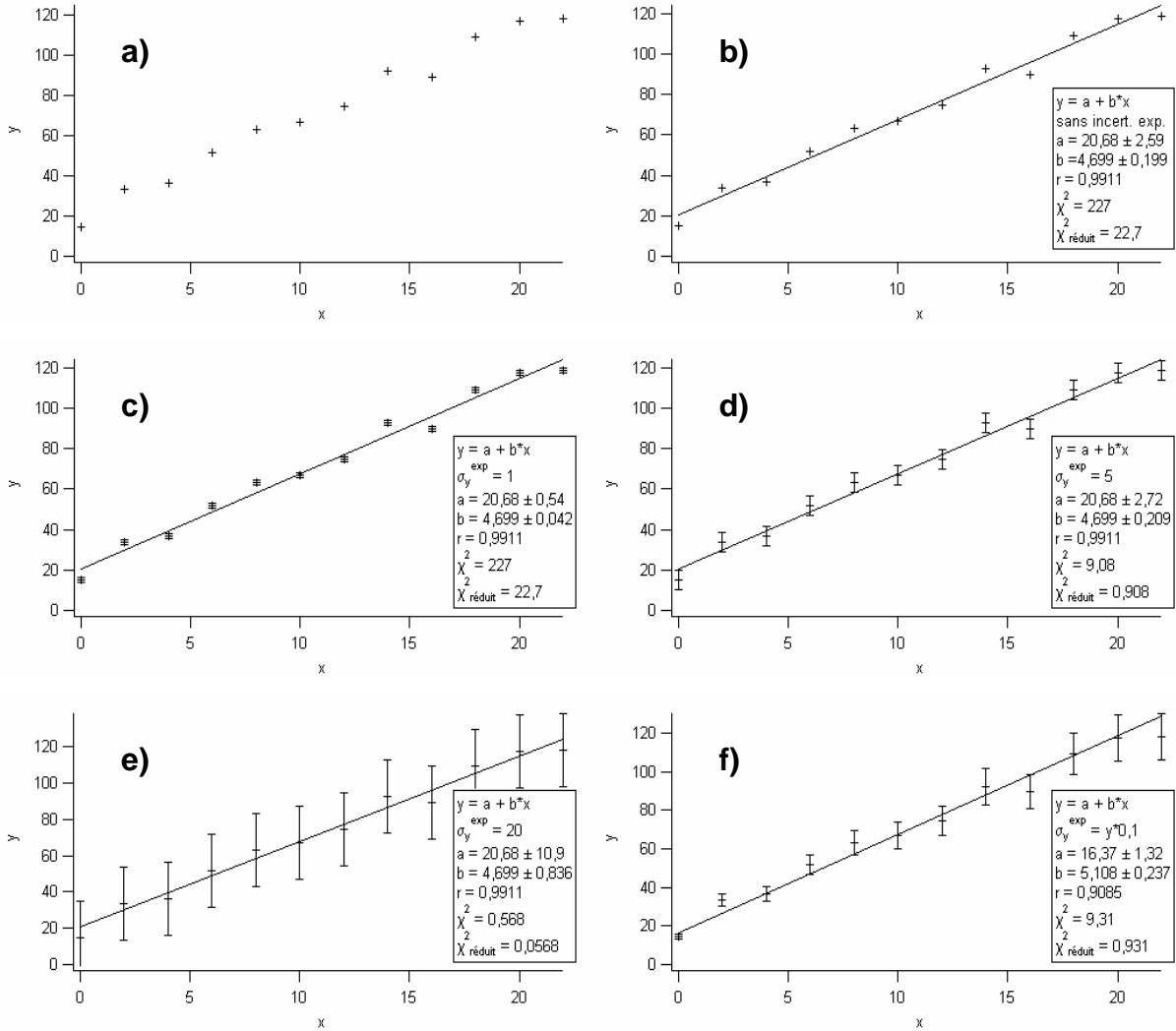
18. - sous Igor : Analysis → Curve Fitting → Fonction = line.

Pour prendre en compte les incertitudes : onglet Data Option → Weighting → nom de la wave dans laquelle on a mis les incertitudes. Laisser par défaut l'option "Standard Dev." (l'autre option sert à la compatibilité avec une ancienne version d'Igor). Attention, la prise en compte des incertitudes est indépendante de l'affichage des barres d'erreur sur le graphe.

- sous Synchronie : Traitement → Modélisation

Synchronie ne permet pas d'afficher les barres d'erreur !

Figure 3 - Régression linéaire : exemples



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

Deux cas différents se présentent :

#### a. Cas où toutes les mesures de $y$ ont la même incertitude-type $\sigma_y^{\text{exp}}$

C'est en particulier ce qui est supposé implicitement si on n'a pas précisé les incertitudes-types  $\sigma_i^{\text{exp}}$ . La méthode utilisée, appelée *méthode des moindres carrés*, consiste à chercher les valeurs de  $a$  et  $b$  qui rendent minimum :

$$\sum_{i=1}^n (y_i - (a + bx_i))^2$$

Cette minimisation conduit directement à :

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

$$b = \frac{n \sum_{i=1}^n x_i y_i - \sum_{i=1}^n x_i \sum_{i=1}^n y_i}{\Delta} \quad (2)$$

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  $n \geq 2$ . D'ailleurs, pour  $n = 1$  elles donnent des formes indéterminées pour  $a$  et  $b$ .

La figure 3.b montre les résultats obtenus pour les données du tableau 3 sans avoir précisé les incertitudes sur  $y$ . Les figures 3.c, d et e présentent les mêmes données avec des incertitudes-types respectives  $\sigma_y^{exp} = 1$ ,  $\sigma_y^{exp} = 5$  et  $\sigma_y^{exp} = 20$ . On peut vérifier sur ces 4 figures que les valeurs de  $a$  et  $b$  obtenues sont identiques (et donc indépendantes de l'incertitude  $\sigma_y^{exp}$ ).

### b. Cas où les mesures de $y$ ont des incertitudes-types différentes

On peut généraliser les résultats ci-dessus au cas où on précise pour chaque mesure  $y_i$  l'incertitude-type  $\sigma_i^{exp}$ . On introduit le poids  $w_i = 1/(\sigma_i^{exp})^2$  de la  $i^{\text{ème}}$  mesure qui est d'autant plus grand que cette mesure est précise. En rendant minimum :

$$\sum_{i=1}^n w_i (y_i - (a + bx_i))^2$$

On obtient :

$$a = \frac{\sum_{i=1}^n w_i x_i^2 \sum_{i=1}^n w_i y_i - \sum_{i=1}^n w_i x_i \sum_{i=1}^n w_i x_i y_i}{\Delta}$$

$$b = \frac{\sum_{i=1}^n w_i \sum_{i=1}^n w_i x_i y_i - \sum_{i=1}^n w_i x_i \sum_{i=1}^n w_i y_i}{\Delta}$$

où

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

La figure 3.f présente les résultats obtenus pour les données du tableau 3 avec  $\sigma_i^{exp} = y/10$ . On constate que les valeurs de  $a$  et  $b$  sont cette fois-ci différentes de celles obtenues avec des incertitudes identiques. En comparant les figures 3.d et 3.f, on voit sur la figure 3.f que la droite passe plus près des points où l'incertitude  $\sigma_i^{exp}$  est plus faible et plus loin de ceux où elle est plus grande. Ceci est cohérent avec la pondération par  $w_i = 1/(\sigma_i^{exp})^2$ .

Le programme de fit fournit également les valeurs des incertitudes-types  $\sigma_a$  et  $\sigma_b$  sur  $a$  et  $b$ . Pour cela, il utilise les résultats du paragraphe suivant.

### 3.1.2 Incertitudes-types $\sigma_a$ et $\sigma_b$ sur les paramètres $a$ et $b$

Pour évaluer  $\sigma_a$  et  $\sigma_b$ , on utilise les formules 1 et 2 et les techniques de propagation des incertitudes indépendantes décrites au paragraphe 2.2.

#### a. Cas où toutes les mesures de $y$ ont la même incertitude

On obtient alors<sup>19</sup> :

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

19. Remarquer que  $\frac{\sigma_a}{\sigma_y}$  et  $\frac{\sigma_b}{\sigma_y}$  ne dépendent que des  $x_i$ .



où :

- si on a fourni l'incertitude-type  $\sigma_y^{exp}$  sur les mesures de  $y$

$$\sigma_y = \sigma_y^{exp}$$

On peut ainsi vérifier sur les figures 3.c, d et e que  $\sigma_a$  et  $\sigma_b$  sont proportionnelles à  $\sigma_y^{exp}$ .<sup>20</sup>

- si on n'a pas indiqué d'incertitudes sur les mesures de  $y$

$$\sigma_y = \sigma_y^{stat}$$

où  $\sigma_y^{stat}$  est l'estimation statistique décrite ci-dessous de l'incertitude-type sur les mesures de  $y$ .

Comme chaque mesure  $y_i$  se distribue autour de la valeur vraie  $a+bx_i$  avec la même incertitude-type  $\sigma_y$ , les écarts  $y_i - (a+bx_i)$  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<sup>21</sup> :

$$\sigma_y^{stat} = \sqrt{\frac{1}{n-2} \sum_{i=1}^n (y_i - (a + bx_i))^2}$$

### b. Cas où les mesures de $y$ ont des incertitudes-types différentes

Les incertitudes-types sur  $a$  et  $b$  sont alors données par :

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

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

A 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 3.d, les incertitudes sont comparables aux écarts à la droite. Les données et leurs incertitudes sont en bon accord avec une loi linéaire<sup>22</sup>.

Sur la figure 3.c, les incertitudes sont en moyenne petites par rapport aux écarts à la droite. Les données et leurs incertitudes ne peuvent être modélisées par une loi linéaire<sup>23</sup>. Soit la loi n'est pas linéaire, soit on a sous-estimé les incertitudes.

---

20. On peut s'étonner que  $\sigma_a$  et  $\sigma_b$  ne dépendent pas des écarts statistiques constatés sur la courbe, en fait on considère que ces écarts sont dus aux incertitudes expérimentales, on ne va donc pas les compter deux fois.

21. Voir *Incertainces et analyse des erreurs dans les mesures physiques*, Paragraphe 8.3, p. 174, J.R. Taylor, Editions Dunod, 2000.

22. Attention, ce n'est pas une preuve de linéarité, c'est juste une compatibilité.

23. Paradoxalement, c'est dans ce cas que les incertitudes sur  $a$  et  $b$  sont les plus faibles. On ne peut donc pas tirer de ces incertitudes un critère de validité de la loi.

Sur la figure 3.e, 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.

Comme on n'a pas précisé les incertitudes, le cas présenté figure 3.b ne permet pas de valider la loi linéaire. Par contre, *si on admet cette loi*, on a l'incertitude statistique sur  $a$  et  $b$ .

**En conclusion :** La discussion qualitative ci-dessus suffit bien souvent. Le paragraphe 3.1.5 propose néanmoins une version quantitative de cette discussion.

### 3.1.4 Coefficient de corrélation linéaire

Dans tous les cas, le programme de fit donne la valeur du coefficient de corrélation linéaire<sup>24, 25</sup> :

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

où  $\bar{x}$  et  $\bar{y}$  sont respectivement les moyennes des  $x_i$  et des  $y_i$ .

$r = 0$  pour un grand nombre de points répartis au hasard et  $r = \pm 1$  pour des points parfaitement alignés.  $r$  est un outil utilisé par les statisticiens 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. Ce n'est pas ici un outil efficace pour discuter de la validité d'une loi linéaire.

### 3.1.5 Le $\chi^2$

On définit le paramètre  $\chi^2$  par<sup>26</sup> :

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

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  $y$ <sup>27, 28</sup>.

On définit ensuite le  $\chi^2$  réduit :

$$\chi_{réduit}^2 = \frac{\chi^2}{n - 2}$$

$n - 2$  est le nombre de degrés de liberté du problème, égal au nombre de points mesurés moins le nombre de paramètres déterminés (ici  $a$  et  $b$ )<sup>29</sup>. Dans le cas où toutes les incertitudes-types sont égales, il est facile d'établir que :

$$\chi_{réduit}^2 = \left( \frac{\sigma_y^{stat}}{\sigma_y^{exp}} \right)^2$$

où  $\sigma_y^{stat}$  est l'incertitude statistique évalué par le logiciel (cf.3.1.2).

24. La formule donnée ci-dessous est valable lorsque tous les  $y$  ont la même incertitude.

25. Sous Igor, il est donné par la variable  $V\_Pr$ . Sous Synchronie, c'est  $r$ .

26. Sous Igor, il est donné par la variable  $V\_chisq$  (pour "chisquare").

27. De manière générale, les paramètres  $a$  et  $b$  donnés par la méthode des moindres carrés sont ceux qui minimisent le  $\chi^2$ .

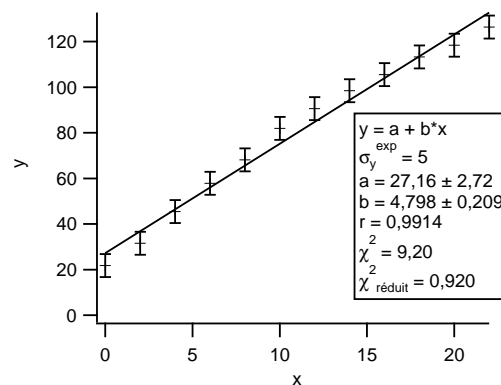
28. Sur les figures 3.c, d et e, on peut vérifier que le  $\chi^2$  est inversement proportionnel au carré de l'incertitude  $\sigma_y^{exp}$ .

29. Dans le cas d'un ajustement par une loi du type  $y = bx$ , on a  $\chi_{réduit}^2 = \chi^2 / (n - 1)$

Le  $\chi_{réduit}^2$  fournit 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  $\chi_{réduit}^2$  est de l'ordre de 1. C'est le cas de la figure 3.d.
- Dans le cas de la figure 3.c,  $\chi_{réduit}^2 \gg 1$  donc  $\sigma_y^{stat} \gg \sigma_y^{exp}$ . La loi n'est pas validée ou  $\sigma_y^{exp}$  a été sous-estimé.
- Le cas  $\chi_{réduit}^2 \ll 1$  correspond à la figure 3.e. Les données peuvent être modélisées par une loi linéaire. Il est cependant probable qu'on a surestimé les incertitudes.

Figure 4 - Loi visiblement non linéaire



Remarque 1 : la figure 4 montre 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. Le tracé des *résidus* peut faciliter l'étude.

Remarque 2 : certains logiciels, comme Igor, renvoient quand même une valeur de  $\chi^2$  même si on ne leur a pas fourni les incertitudes  $\sigma_i^{exp}$  (voir figure 3.b). Cette valeur est calculée en prenant  $\sigma_i^{exp} = 1$ . On ne peut alors pas tirer grand chose du  $\chi_{réduit}^2$  si ce n'est la valeur de  $\sigma_y^{stat}$  (on a alors  $\chi_{réduit}^2 = (\sigma_y^{stat})^2$ ).

### 3.2 Cas général d'un ajustement par une loi $y = f(x)$

Les principes développés dans le cas de la régression linéaire s'appliquent encore dans le cas d'un ajustement par une fonction quelconque  $y = f(x)$ , dépendant d'un certain nombre de paramètres. Les logiciels d'ajustement déterminent numériquement les valeurs des paramètres qui minimisent le  $\chi^2$ .

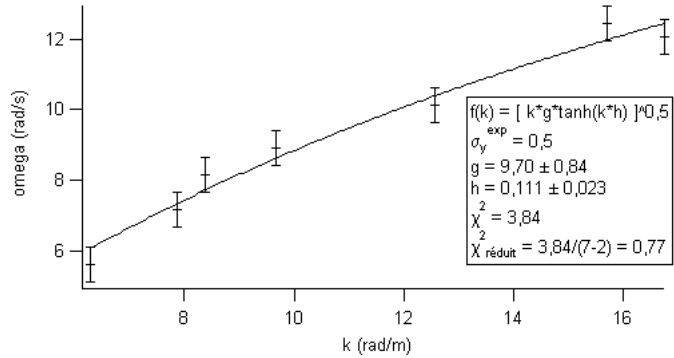
Remarque : Contrairement au cas linéaire, il faut ici donner des valeurs initiales aux paramètres de la fonction. Ceci doit être fait avec une précision suffisante pour que la procédure de minimisation converge.

#### En cas d'échec de la procédure :

- contrôler que le fit est fait sur les bonnes variables en ordonnée et abscisse.
- 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  $y = a(x - c)^b + d$ .
- 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.
- 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 fitter en effectuant une sélection sur la courbe.

**Figure 5 - Relation de dispersion des ondes de gravité**

A titre d'exemple, la figure 5 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 :  
 $f(k) = (k * g * \tanh(k * h))^{1/2}$   
( $g$  est l'intensité du champ de pesanteur).

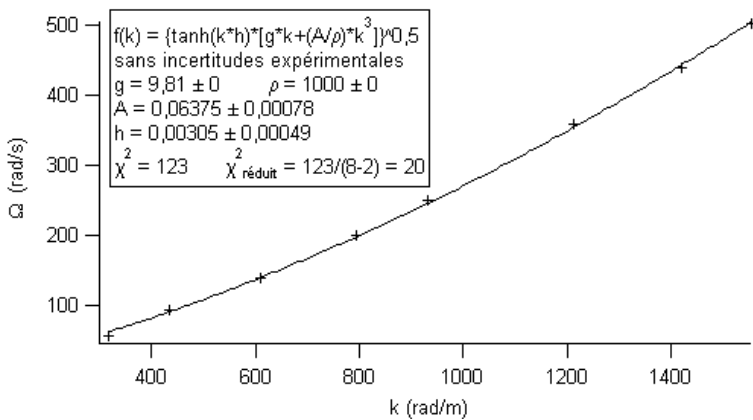


Visuellement, on constate que les données et les incertitudes qui leur sont associées sont en accord avec la loi proposée. On obtient une détermination des paramètres  $g$  et  $h$  avec leurs incertitudes. Par ailleurs, le logiciel évalue le  $\chi^2$ . On peut en déduire le  $\chi^2_{réduit}$  (on a 7 points de mesures et 2 paramètres à déterminer donc 5 degrés de liberté), qui confirme quantitativement l'analyse visuelle.

De manière similaire, la figure 6 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  $f(k) = ((k * g + (A/\rho) * k^3) * \tanh(k * h))^{1/2}$  ( $A$  et  $\rho$  sont respectivement la tension superficielle et la masse volumique de l'eau). On a choisi de bloquer les paramètres  $g$  et  $\rho$  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.

**Figure 6 - Relation de dispersion des ondes à la surface de l'eau**



Ici, on n'a pas rentré d'incertitude-type expérimentale  $\sigma_y^{exp}$ . 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.

## Annexe : Rappels élémentaires de statistique

### A.1 Distribution de probabilité d'une grandeur aléatoire

On appelle distribution de probabilité d'une grandeur aléatoire  $x$ <sup>30</sup> à valeurs continues la fonction  $f(x)$  telle que la probabilité pour que  $x$  soit comprise entre  $a$  et  $b$  s'écrive :

$$P(a < x < b) = \int_a^b f(x)dx$$

$f(x)$  est donc une densité de probabilité :  $f(x)dx$  est la probabilité que la variable aléatoire  $x$  prenne une valeur comprise entre  $x$  et  $x + dx$ . On en déduit les deux propriétés suivantes :

$$f(x) \geq 0$$

$$\int_{-\infty}^{+\infty} f(x)dx = 1$$

### A.2 Valeur moyenne et écart-type

La valeur moyenne (ou espérance)  $X$  d'une grandeur aléatoire  $x$  à valeur continue est donnée par :

$$X = \int_{-\infty}^{+\infty} xf(x)dx$$

L'écart-type  $\sigma$  de la distribution des valeurs de  $X$  est défini par :

$$\sigma^2 = \int_{-\infty}^{+\infty} (x - X)^2 f(x)dx$$

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 l'écart-type est faible.

$\sigma^2$  est appelée variance de la distribution.

Cas d'une distribution de probabilité discrète : quand la grandeur aléatoire ne prend que des valeurs discrètes  $x_i$  avec des probabilités respectives  $p_i$ , la valeur moyenne et l'écart-type de la distribution sont respectivement donnés par :  $X = \sum_i x_i p_i$  et  $\sigma^2 = \sum_i (x_i - X)^2 p_i$

### A.3 Exemples de distributions de probabilité

#### – La distribution de Gauss ou Loi Normale

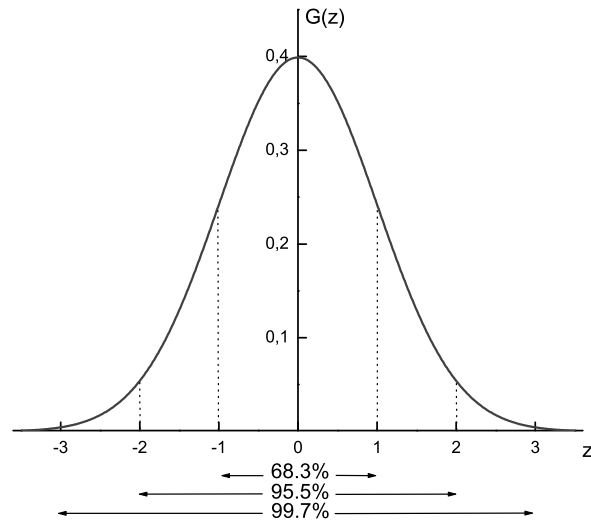
La distribution de Gauss, de valeur moyenne  $X$  et d'écart-type  $\sigma$  s'écrit :

$$G_{X,\sigma}(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-(x-X)^2/2\sigma^2}$$

---

30. Pour simplifier les notations, on désigne par la même lettre  $x$  la variable aléatoire et sa valeur.

**Figure 7** - La distribution de Gauss ou Loi Normale : c'est une distribution symétrique autour de la moyenne dont le graphe est présenté ci-contre en fonction de  $z = (x - X)/\sigma$ . Comme indiqué sur la figure et dans le tableau 4, la probabilité que la valeur de  $x$  soit à moins d'un écart-type de la moyenne (en valeur absolue) est de 68 %. Cette probabilité passe à 99.7 % pour un écart à la moyenne de 3 écarts-types. L'importance physique de la distribution gaussienne est discutée au paragraphe A.4



$\alpha$	0	0,5	1	1,5	2	2,5	3	3,5	4
<b>P(%)</b>	0	38	68	87	95,4	98,8	99,7	99,95	99,99

**Table 4** - Loi Normale : probabilité  $P$  qu'une mesure de  $x$  se trouve dans l'intervalle  $[X - \alpha\sigma, X + \alpha\sigma]$

#### – La 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 :

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

Cette distribution de probabilité, définie uniquement pour les valeurs entières  $k$  a pour moyenne  $np$  et pour écart-type  $\sqrt{np(1 - p)}$ . Elle n'est pas symétrique autour de sa moyenne. 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  $np$  et d'écart-type  $\sqrt{np(1 - p)}$  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.<sup>31</sup>

#### – La 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. 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 :

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

$N$  est le nombre moyen d'évènements observés.

On rencontre par exemple la distribution de Poisson lorsqu'on compte, pendant un temps donné, un nombre de désexcitations radioactives, ou le nombre de photons d'un faisceau laser. Elle décrit également le nombre de molécules se trouvant à l'intérieur d'un petit volume de gaz parfait.

31. Pour une analyse plus détaillée de cette approximation voir : *Physique Statistique*, complément I.C, p.62, B. Diu, C. Guthmann, D. Lederer, B. Roulet, éditions Hermann, 1995.

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

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  $p$  (dans la pratique pour  $n > 50$  et  $N < 5$ ). Comme la loi binômiale, la distribution de Poisson est très proche d'une Gaussienne dès que le nombre d'évènements observés est grand.

#### A.4 Importance physique de la Loi Normale

Dans la pratique, si 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 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 (la version mathématique de ce théorème est appelée Théorème Central Limite)<sup>32</sup>.

---

<sup>32</sup>. On trouvera une illustration physique simple de ce théorème dans : *Incertitudes et analyse des erreurs dans les mesures physiques*, Paragraphe 10.5, p. 235, J.R. Taylor, Editions Dunod, 2000.