

Méthodes 2 : Mesures et incertitudes

Plan du cours

I Processus de mesure : variabilité des résultats et incertitude-type	2
I.1 Un exemple pour poser le problème	2
I.2 Variabilité des résultats	2
I.3 Erreur et incertitude	3
I.4 Résultat d'un mesurage	4
II Évaluation de l'incertitude-type	4
II.1 Série de mesure : Type A	5
II.2 Mesure unique : Type B	7
II.3 Mesure indirecte : Incertitude-type composée	10
III Validité d'un résultat expérimental	11
III.1 Compatibilité entre 2 valeurs	11
III.2 Vérification d'une loi : modélisation linéaire	12
IV Méthode Monte-Carlo	14
IV.1 Incertitude-type composée avec MC	14
IV.2 Modélisation affine avec MC	15

À savoir



Faire la distinction entre erreur de mesure et incertitude de mesure

I.1,3

Connaître les 2 types d'erreurs, leurs causes, leurs effets

I.3

Signification de l'écart-type pour la loi normale

I.4

Connaître les 2 approches (statistique → type A, et probabiliste → type B) pour évaluer une incertitude-type

II.1,2

Formule de l'incertitude-type pour le type A et pour le type B

II.1,2

Formules de l'incertitude-type composée

II.3

Formules de l'écart normalisé et du z-score + critère de validation

III.1

À savoir faire



Calculer l'incertitude de type A pour une série de valeurs expérimentales.

II.1

Calculer l'incertitude de type B pour une mesure expérimentale unique.

II.2

Lister les sources d'incertitude possibles

I.2

Tester la validité d'un résultat expérimental.

III.1

Réaliser une régression linéaire et tester la validité de ce modèle

III.2

Réaliser une simulation de Monte-Carlo pour simuler une série de valeurs et déterminer une incertitude-type composée.

IV.1

Réaliser une simulation de Monte-Carlo pour déterminer les incertitudes-types sur les paramètres d'un modèle linéaire.

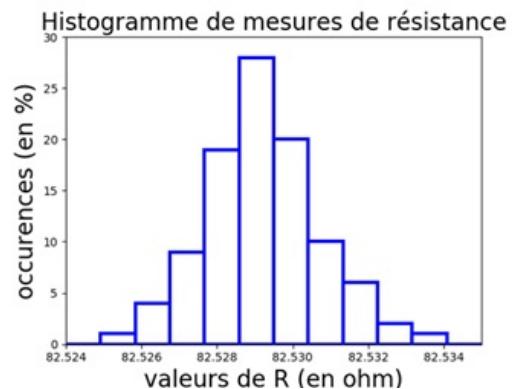
IV.2

I Processus de mesure : variabilité des résultats et incertitude-type

I.1 Un exemple pour poser le problème

Expérience 1 :

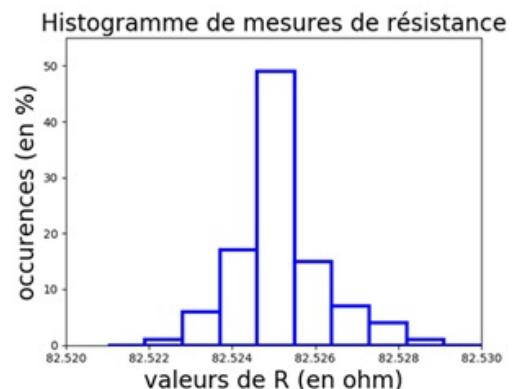
On a mesuré la valeur de la résistance d'un conducteur ohmique 2000 fois, puis représenté les résultats obtenus sur l'histogramme ci-contre. Que peut-on dire sur la répartition des résultats ?



Comment expliquer une telle variabilité des résultats ?

Expérience 2 :

On suit le même protocole que dans l'expérience 1, mais avec un autre multimètre, et on obtient l'histogramme ci-contre. Les résultats sont-ils identiques ?

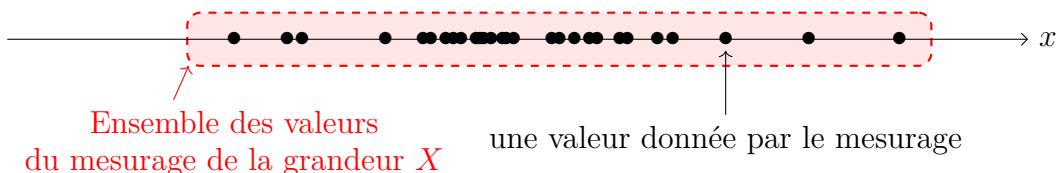


Quelle est LA valeur de cette résistance ? Toute la problématique de la détermination de la mesure d'une grandeur est là :

- Quel résultat choisir ?
- Comment estimer sa « précision » ?
- Comment comparer à une valeur de référence ?

I.2 Variabilité des résultats

En sciences expérimentales, le mesurage d'une grandeur physique est un processus complexe, et la répétition de la mesure conduit naturellement à une dispersion des valeurs observées. **Cette variabilité fait partie du mesurage !**



Plusieurs raisons peuvent expliquer cette variabilité :

- La grandeur à mesurer n'est pas parfaitement définie. Exemples : la largeur d'une table peut varier suivant la longueur, la surface d'un liquide n'est pas plane, etc.
- Les conditions environnementales (température, pression, etc.) ne sont pas parfaitement stables.
- L'instrument de mesure est source d'erreur (temps de réponse, exactitude, sensibilité).
- L'opérateur ne refait jamais la même mesure exactement dans les mêmes conditions (fatigue, erreurs de parallaxe, effet de ménisque dans une pipette, etc.)



Remarque

Une mesure comporte en général plusieurs opérations dont chacune peut être source de variabilité. Il sera important de savoir distinguer les sources de variabilité importante de celles qui sont négligeables.

I.3 Erreur et incertitude



Vocabulaire

Erreur de mesure : c'est la différence entre la valeur mesurée et la valeur vraie

$$\text{erreur} = |x_{\text{mes}} - x_{\text{vraie}}|$$

Elle possède deux composantes distinctes :

- L'erreur aléatoire : elle provient des variations temporelles et spatiales non prévisibles de grandeurs d'influence, elle n'a pas d'effet sur la valeur moyenne.
- L'erreur systématique : la mesure est systématiquement décalée par rapport à la valeur vraie. Elle provient d'un effet reconnu d'une grandeur d'influence (erreur de protocole, défaut d'un appareil, etc.).

Fidélité et justesse : Une mesure est d'autant plus **fidèle** que l'erreur aléatoire est faible, d'autant plus **juste** que l'erreur systématique est faible.

Incertitude de mesure : c'est un paramètre qui caractérise la dispersion des valeurs autour de la valeur moyenne des différentes valeurs mesurées d'un mesurande. Elle peut être vue comme un « doute » sur la valeur mesurée.

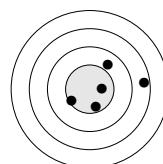
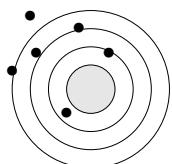
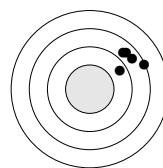
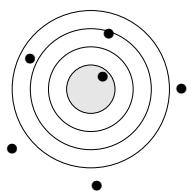


Remarques

- Il ne faut pas confondre erreur et incertitude : une grande incertitude implique une plus grande chance d'erreur, mais une faible erreur n'implique pas une faible incertitude, ça peut être un coup de chance !
- En augmentant le nombre de mesurages, on peut réduire les effets de l'erreur aléatoire sur la précision du résultat, mais on ne réduit pas l'effet de l'erreur systématique.
- Dans le cas d'une erreur systématique non négligeable devant la précision requise, on peut appliquer une correction au résultat.

Application

On a représenté les résultats d'un mesurage par une analogie avec une cible de tir à l'arc (le centre de la cible représente la valeur varie). Préciser dans chaque cas si la mesure est juste et/ou fidèle ainsi que le(s) type(s) d'erreur présent(s).



I.4 Résultat : meilleur estimateur et incertitude-type associée à un mesurage

Donner le résultat du mesurage d'une grandeur physique X , c'est :

- Donner la meilleure estimation de sa valeur vraie (souvent inconnue, ce qui revient à une absence de cible dans les schémas de la page 3 !)

ET

- Caractériser la dispersion des valeurs que peut prendre cette grandeur physique. Du fait de cette variabilité intrinsèque au processus mesurage, on fait appel à la théorie des probabilités pour décrire correctement la mesure d'une grandeur, en utilisant la notion d'écart-type que nous expliciterons plus loin.



Méthode

Résultat d'un mesurage : Le résultat d'un mesurage n'est pas une valeur unique mais un ensemble de valeurs numériques, raisonnablement attribuables au mesurande.

Notation : Tout résultat expérimental doit être accompagné d'une **incertitude de mesure** et d'une **unité** :

$$X = x_{\text{mes}} \text{ unité} \quad \text{avec} \quad u(X) = \dots \text{ unité}$$

avec : x_{mes} = meilleure estimation de la grandeur mesurée (valeur expérimentale dans le cas d'une mesure unique, ou moyenne arithmétique des valeurs expérimentales)

$u(X)$ = incertitude-type sur la mesure de la grandeur X , qui définit un intervalle de confiance pour l'estimation de la grandeur mesurée, elle est liée à la notion d'écart-type



Remarques

- On écrit l'incertitude-type avec **2 chiffres significatifs** (arrondi au supérieur), et le chiffre significatif de x_{mes} doit coïncider avec le dernier chiffre significatif de $u(X)$.
- Ce n'est pas la résolution de l'instrument de mesure qui fixe le nombre de chiffres significatifs de la valeur mesurée.

Application

Les résultats suivants sont-ils écrits correctement ? Si non, les corriger.

$$R = 18,3 \Omega \text{ avec } u(R) = 0,21 \Omega$$

$$R = 11,34 \Omega \text{ avec } u(R) = 0,16 \Omega$$

$$R = 16 \Omega \text{ avec } u(R) = 0,102 \Omega$$

$$R = 18,367 \Omega \text{ avec } u(R) = 0,20 \Omega$$

II Différents modes d'évaluation de l'incertitude-type sur une grandeur

On rencontre deux situations types selon ce qui limite la précision du résultat obtenu :

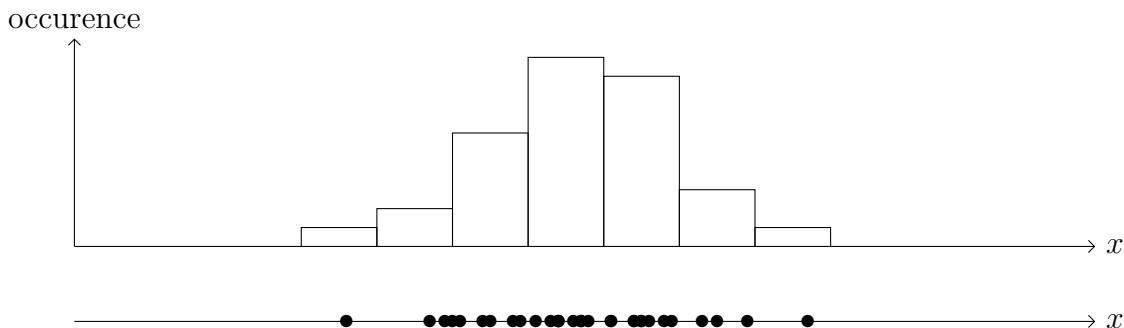
- Cas n°1 : La précision du résultat est limitée par la répétabilité du protocole expérimental (chaque observation donne des résultats différents). On réalise alors un traitement statistique de l'incertitude : évaluation de **type A**.
- Cas n°2 : Certaines expériences n'ont pas de variabilité observée (en reproduisant la mesure, on retrouve systématiquement le même résultat), ou il n'est pas possible de reproduire plusieurs fois le protocole de mesurage. Il faut alors estimer théoriquement la variabilité de la mesure sans l'observer, on réalise un traitement probabiliste de l'incertitude : évaluation de **type B**.

II.1 Série de mesures : évaluation de type A de l'incertitude-type

Si on place sur un axe un ensemble $\{x_1, \dots, x_N\}$ de N mesures de la valeur d'une grandeur X , comment donner le résultat ?



Un outil pratique pour visualiser la dispersion des observations réalisées est l'histogramme, grâce auquel on voit que la répartition des valeurs mesurées suit très souvent une courbe « en cloche ».

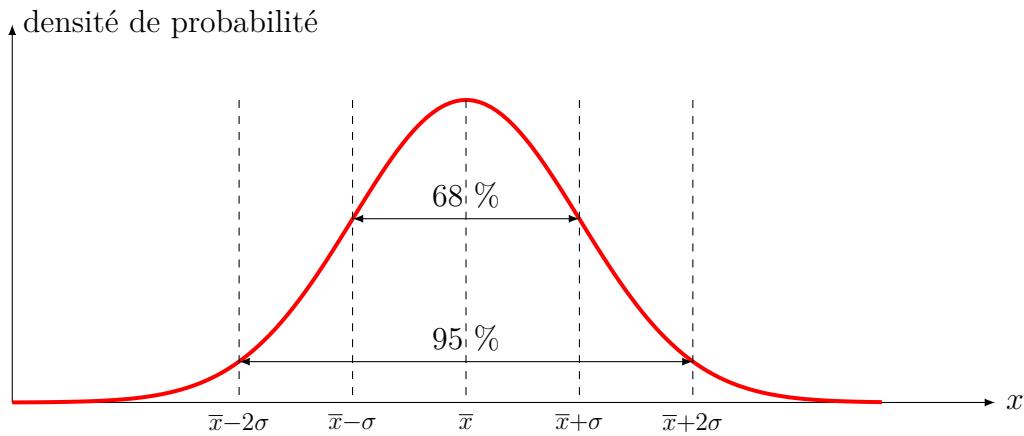


Cet histogramme « en cloche » peut être assimilé à une fonction « gaussienne », on dit que x suit une loi normale (c'est la loi de probabilité la plus utilisée en statistique).

Propriétés

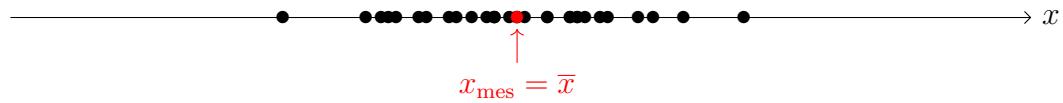
Propriétés d'une loi normale :

- Elle est centrée par rapport à la moyenne, notée \bar{x} , des valeurs prises par la variable x .
- Son étalement est défini par l'écart-type σ : la probabilité d'obtenir une valeur dans l'intervalle $[\bar{x} - \sigma ; \bar{x} + \sigma]$ est de 68% .



Lors d'une série de mesures, en l'absence d'erreur systématique :

- Il n'y a pas de raison pour que les résultats se répartissent plus d'un côté que de l'autre de la moyenne arithmétique des résultats, donc la meilleure estimation de la valeur x_{vraie} est obtenue en faisant la moyenne des valeurs mesurées $x_{mes} = \bar{x}$



- La dispersion des valeurs lors d'une observation est donnée par l'écart-type des mesures (écart-type de l'échantillon) :

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



Série de mesure → Type A (approche statistique)

Si une grandeur X est estimée à partir de n observations répétées indépendantes $\{x_1, \dots, x_n\}$, et si $\sigma(X)$ est l'écart-type expérimental d'une mesure (obtenu à partir de ces n valeurs), alors :

- Sa meilleure estimation est donnée par la moyenne des n valeurs :

$$X = \bar{x} = \frac{x_1 + x_2 + \dots + x_n}{n}$$

- L'incertitude-type $u(X)$ sur son estimation \bar{x} vaut l'écart-type de la moyenne :

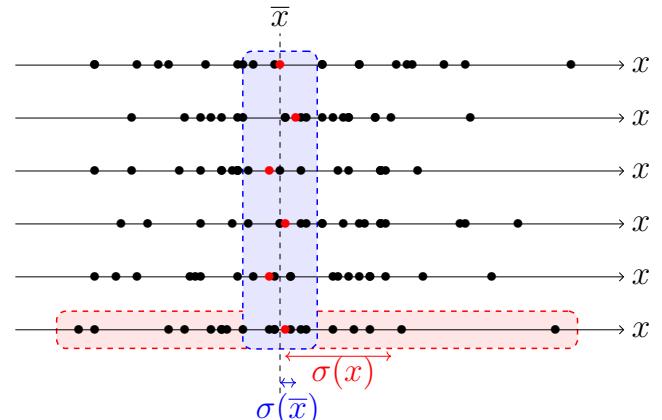
$$u(X) = \sigma(\bar{x}) = \frac{\sigma}{\sqrt{n}}$$

avec $\sigma(x) = \text{écart-type de l'échantillon} = \sqrt{\frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})^2}$



Remarques

- $u(X)$ dépend de la nature du protocole expérimental mis en œuvre, du matériel utilisé, des conditions expérimentales et de l'habileté des expérimentateurs.
- Pour utiliser une incertitude de type A, il faut au minimum 5 à 10 mesures, mais c'est inutile de dépasser 20 mesures (une augmentation de n diminue l'incertitude-type mais cette diminution n'est pratiquement plus efficace au delà de $n = 10$ mesures).
- L'écart-type de l'échantillon $\sigma(x)$ est donné dans le menu statistique de la calculatrice (« S_x » sur les TI, « sx » sur les Casio, « écart-type échantillon » sur les Numworks).
- Justification qualitative du facteur $\frac{1}{\sqrt{n}}$:
Si on effectue n mesures lors du mesurage, \bar{x} est une moyenne parmi d'autres possibles : si on refaisait à nouveau n mesures, on obtiendrait une autre moyenne. Or la distribution de l'ensemble des moyennes est bien moins dispersée que l'ensemble des mesures uniques (dispersion divisée par \sqrt{n}), l'estimation à partir d'une moyenne est « meilleure » que sur une mesure seule.



Application

On effectue 10 mesures de la résistance R d'un conducteur ohmique, les résultats sont présentés dans le tableau ci-dessous (valeurs en Ω) :

523,6	523,5	523,4	523,5	523,6	523,4	523,5	523,5	523,6	523,4
-------	-------	-------	-------	-------	-------	-------	-------	-------	-------

Q1. Déterminer (avec le menu « statistiques » de la calculatrice) :

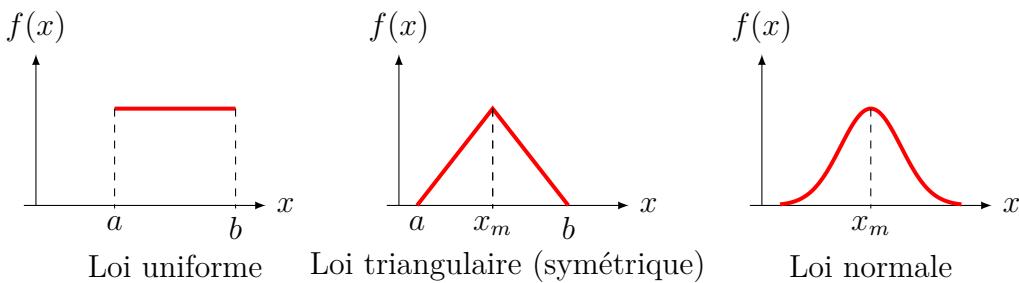
- La valeur moyenne de R :
- L'écart-type expérimental de la série de mesures :
- L'écart-type sur la moyenne :

Q2. Écrire correctement le résultat du mesurage :

II.2 Mesure unique : évaluation de type B de l'incertitude-type

Dans le cas d'une mesure x isolée (ce qui est le plus fréquent en TP, la plupart du temps il n'est pas possible de répéter la mesure et il faut se contenter d'un résultat unique!), on utilise aussi la théorie des probabilités déterminer l'incertitude-type sur la mesure. Il existe différents modèles de densités de probabilité :

- La loi uniforme sur un intervalle $[a, b]$
- La loi triangulaire sur un intervalle $[a, b]$ et de maximum en x_m .
- La loi Gaussienne (encore appelée loi normale) centrée sur une valeur x_m et d'écart-type σ



L'incertitude-type est alors donné par l'écart-type de la distribution retenue.

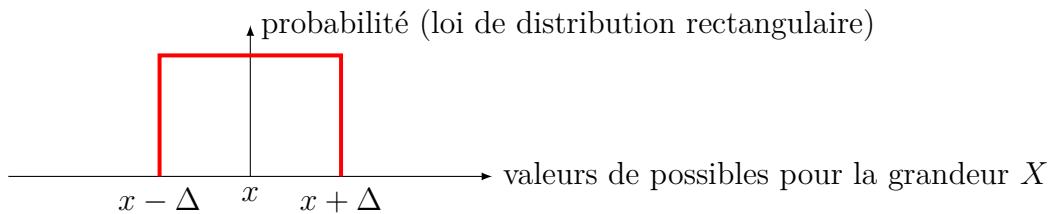
Dans le cas d'une mesure unique, on suppose que la densité de probabilité est à support borné (type uniforme ou triangulaire symétrique) et on évalue expérimentalement l'intervalle-support $[x - \Delta, x + \Delta]$ dans lequel on estime qu'on aurait 100% de chances d'obtenir n'importe le résultat de mesure si on répétait l'expérience un grand nombre de fois.

Pour choisir parmi ces différents modèles bornés, on procède à l'évaluation de l'erreur possible (souvent donnée par la précision de l'appareil de mesure, notée Δ), et on se questionne sur la distribution de probabilité en dessous de cette précision : la valeur vraie a-t-elle plus de chance de se trouver au centre de l'intervalle ou pas ?

Par habitude, n'ayant aucune information en dessous de la précision Δ , on estime que la valeur vraie x_{vraie} a « autant de chances » de prendre n'importe quelle valeur entre $x - \Delta$ et $x + \Delta$ donc on considère une probabilité égale pour toutes les valeurs comprises entre $x - \Delta$ et $x + \Delta$, donc on privilégie la distribution uniforme, et l'incertitude-type sera évaluée avec la loi rectangulaire.

❤️ Mesure unique → Type B (approche probabiliste)

L'incertitude-type dans le cas d'une loi rectangulaire est liée à la précision Δ par : $u(X) = \frac{\Delta}{\sqrt{3}}$



💡 Remarque

Si on considère une distribution triangulaire symétrique, l'incertitude-type de la mesure (donné par l'écart-type de la distribution) vaut : $u(X) = \frac{\Delta}{\sqrt{6}}$.



Estimation de la précision des appareils de mesure (notée Δ)

- La précision Δ des **instruments de mesure gradués** est égale à une **demi-graduation**.
- La précision d'un **multimètre** est donnée par le constructeur de l'appareil : « $n\%L + n \text{ dgt}$ », ou « $n\%L + n\text{UR}$ », ce qui signifie : $n\%$ de la valeur lue + n digit 1 digit (dgt), ou 1 Unité de Représentation (UR) correspond à l'unité du dernier chiffre affiché à droite de l'écran, c'est-à-dire la plus petite valeur qui puisse être affichée sur le calibre choisi.
- **En chimie**, la précision est donnée par la tolérance inscrite sur la verrerie (exemple : fiole jaugée de $500 \text{ mL} \pm 0,25 \text{ mL}$ donc $\Delta = 0,25 \text{ mL}$).
- Dans le cas où la source d'erreur prépondérante n'est **pas liées à un appareil de mesure**, la précision Δ est estimée de façon réaliste, en prenant par exemple la demi-étendue de la plage de valeurs sur laquelle on est sûr d'avoir la valeur vraie.



Remarques

- La demi-étendue Δ doit être prise la plus faible possible selon les critères personnels de l'expérimentateur et selon les conditions de l'expérience, il faut analyser la situation : par exemple avec un instrument gradué, si la valeur lue tombe « pile » sur une graduation, on peut réduire Δ à $\frac{1}{4}$ de graduation, par contre elle tombe entre deux graduations de l'instrument, on choisira $\Delta = \frac{1}{2}$ graduation, ou si l'expérimentateur est très peu sûr de lui, il peut prendre $\Delta = 1$ graduation. Dans tous les cas, justifier le choix fait.
- Pour un appareil à lecture double sur une échelle graduée (règle, balance avec tare, etc.) : il y a deux causes d'erreur (lors de la mesure du zéro ou du tarage et lors de la lecture finale). L'erreur due à la résolution va se produire deux fois mais il est probable qu'il y ait une compensation partielle des erreurs : il faut alors multiplier le demi-intervalle sur une mesure par $\sqrt{2}$ (conséquence de la propagation des incertitudes dans le cas d'une soustraction, voir page 10).
- Lorsque lors d'une mesure, il peut y avoir plusieurs sources d'incertitude :
 - les erreurs de repérage par un expérimentateur (netteté d'une image, parallaxe, maximum d'un signal à l'oscilloscope etc.), l'incertitude associée est notée $u_{\text{repérage}}$
 - les erreurs de lecture (liées aux graduations sur une règle, un rapporteur), l'incertitude associée est notée u_{lecture}
 - les erreurs liées à un appareil, l'incertitude associée est notée u_{appareil}
 Il faudrait, pour être rigoureux, tenir compte de toutes et calculer :

$$u(X) = \sqrt{u_{\text{repérage}}^2 + u_{\text{lecture}}^2 + u_{\text{appareil}}^2}$$

Cependant, pour simplifier, nous ne garderons que celle dont l'influence est la plus grande (en général le repérage par l'expérimentateur).

- Dans le cas où l'on a procédé à une évaluation de type A et de type B, on admet que l'incertitude-type globale s'écrit :

$$u(X) = \sqrt{u_A^2(X) + u_B^2(X)}$$

Applications :

Déterminer dans chaque mesurage la valeur et l'incertitude-type liée à la lecture ou à l'appareil, puis écrire le résultat de la mesure :



Thermomètre



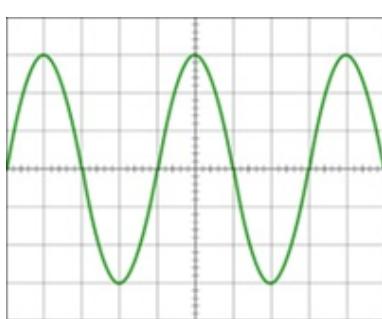
Baromètre (unité mbar)



Fiole jaugée (pleine)



Multimètre utilisé en ampèremètre (précision : $0,5\%L + 0,5$ dgt)



Oscilloscope (mesure de T) calibre 10 ms/DIV



Règle

II.3 Mesure indirecte : évaluation de l'incertitude-type composée

On cherche à obtenir la valeur z d'une grandeur Z à partir de la mesure effective d'autres grandeurs (X et Y) dont elle dépend.

On connaît la relation théorique entre les diverses grandeurs $Z(X, Y)$. L'incertitude sur z est directement reliée à l'incertitude sur les mesures x et y .



Incertitude-type composée

Relation	$Z = X + Y$	$Z = X - Y$	$Z = X \times Y$	$Z = \frac{X}{Y}$
Incertitude-type composée	$u(Z) = \sqrt{u(X)^2 + u(Y)^2}$	$\frac{u(Z)}{ z } = \sqrt{\left(\frac{u(X)}{x}\right)^2 + \left(\frac{u(Y)}{y}\right)^2}$		



Remarques

- Les formules pour le produit simple ($Z = X \times Y$ ou $Z = \frac{X}{Y}$) sont des cas particuliers d'une formule plus générale où Z est un produit de puissances : $Z = X^\alpha Y^\beta$, pour laquelle on a :
- $$\frac{u(Z)}{|z|} = \sqrt{\alpha^2 \left(\frac{u(X)}{x}\right)^2 + \beta^2 \left(\frac{u(Y)}{y}\right)^2}$$
- Dans les cas où la relation $Z(X, Y)$ est plus compliquée que celles données ci-dessus, l'énoncé vous fournira la formule reliant l'incertitude sur z aux incertitudes de mesure (ou le logiciel Gum_MC : http://jeanmarie.biansan.free.fr/gum_mc.html).

Applications :

- Q1. On souhaite déterminer la vitesse des ultrasons dans l'air à l'aide de la relation $c = \lambda \times f$.

On a obtenu par mesure directe :

$\lambda = 8,3 \text{ mm}$ avec $u(\lambda) = 0,6 \text{ mm}$ et $f = 4,0000 \times 10^4 \text{ Hz}$ avec $u(f) = 58 \text{ Hz}$.

Donner le résultat de l'estimation de c .

- Q2. Justifier le facteur $\sqrt{2}$ dans l'évaluation de l'incertitude-type d'une longueur mesurée à la règle.

III Validité d'un résultat expérimental

III.1 Compatibilité entre 2 valeurs

Pour valider le résultat d'un mesurage, on peut comparer la valeur obtenue à la valeur obtenue par un autre procédé de mesurage (ou par une autre personne), ou à une valeur de référence donnée dans des tables. Pour cela, on dispose de deux outils : l'écart normalisé et z-score.



Écart normalisé et z-score

Écart normalisé E_N :

Pour tester la compatibilité entre 2 valeurs mesurées x_1 et x_2 assorties des incertitudes-types u_1 et u_2 :

$$E_N = \frac{|x_1 - x_2|}{\sqrt{u_1^2 + u_2^2}}$$

Z-score z :

Pour tester la validité entre une valeur mesurée x_{mes} assortie de l'incertitude-type $u(X)$ et une valeur de référence x_{ref} :

$$z = \frac{|x_{\text{mes}} - x_{\text{ref}}|}{u(X)}$$

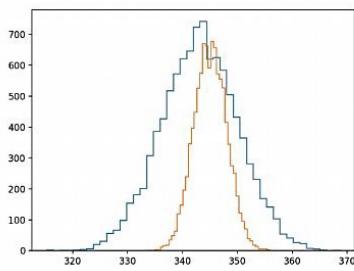
Critère : 2 valeurs sont compatibles si $E_N < 2$ ou $z < 2$.



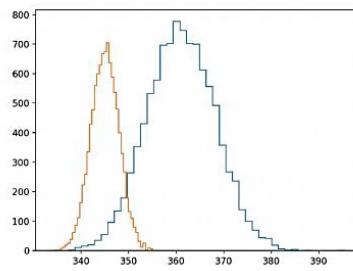
Remarques

- Le choix du seuil est arbitraire mais utilisé dans de nombreux domaines. Il est lié au fait que l'intervalle $[x_{\text{mes}} - 2u(X); x_{\text{mes}} + 2u(X)]$ est un intervalle de confiance à 95% pour une distribution qui suit une loi normale (voir courbe page 5). Si $z \ll 2$, la compatibilité de x_{mes} et x_{ref} est assurée avec une probabilité d'environ 95%.
- Si l'écart normalisé est supérieur à 2, les résultats ne sont pas compatibles. Il faut chercher les origines de cette incompatibilité : variabilité des mesures, erreur de manipulation de calcul, sous-estimation des incertitudes, etc.

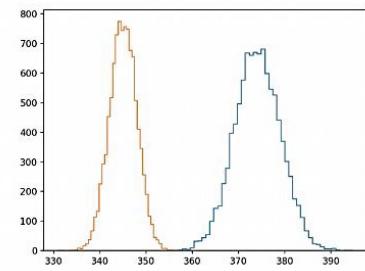
Les 3 graphiques ci-dessous représentent des distributions avec en écart normalisé z :



$$z = 0,3$$



$$z = 2,1$$



$$z = 5,0$$

Application :

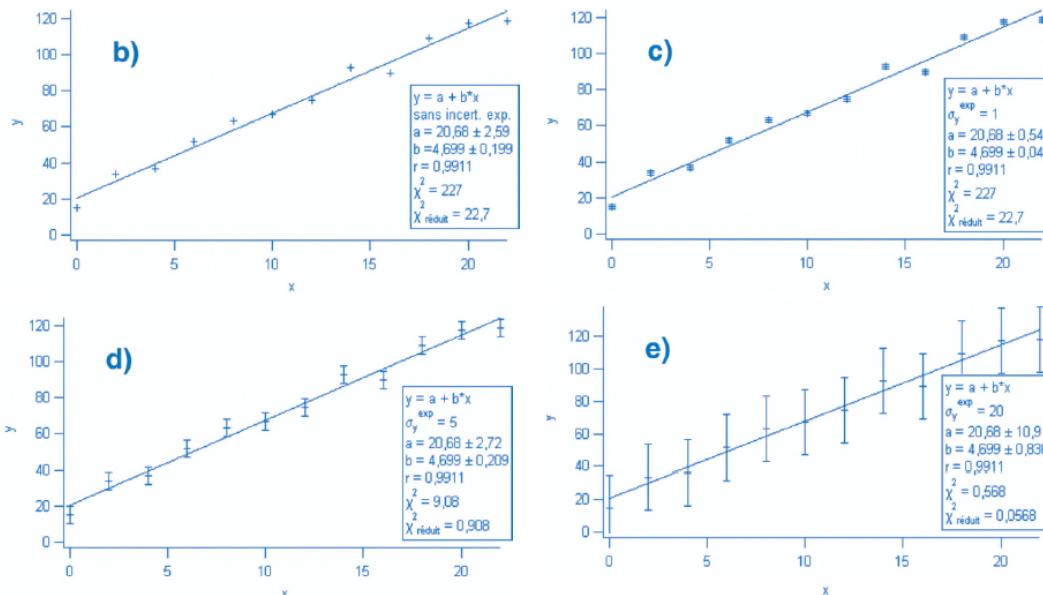
La valeur obtenue pour la célérité des ultrasons dans l'application précédente est-elle en accord avec la valeur tabulée (prise à la même température) : $c_{\text{tab}} = 340 \text{ m}\cdot\text{s}^{-1}$?

III.2 Vérification d'une loi : modélisation linéaire

La vérification d'une loi passe souvent par le tracé d'un graphe, on cherche alors à **représenter une droite afin de conclure facilement sur la validité (ou non) de la loi.**

Coefficient de corrélation :

On a représenté ci-dessous les mêmes mesures expérimentales, mais avec des incertitudes différentes. Pour vérifier si la loi étudiée expérimentalement est affine, on effectue une régression linéaire.



Pour caractériser l'alignement des points expérimentaux, la calculatrice ou le logiciel (Excel/Openoffice) propose un coefficient appelé coefficient de corrélation linéaire r (ou r^2). Il est égal à 1 si la relation est linéaire et s'en écarte si ce n'est pas le cas. Cependant ce coefficient ne permet pas de déterminer si l'écart à 1 est dû aux incertitudes de mesure ou si cet écart est dû au fait que la relation n'est pas linéaire. L'exemple ci-dessus montre que ce coefficient est le même quelles que soient les incertitudes. Si l'on peut croire à la pertinence de l'ajustement dans le cas d), il est plus difficile de s'en convaincre dans les autres cas.

⚠ Le coefficient de corrélation n'est donc pas pertinent pour vérifier une relation linéaire entre les données.

L'existence d'une droite moyenne ne suffit pas à conclure sur la linéarité de la relation. Il faut en plus étudier la répartition des points expérimentaux par rapport à la droite moyenne, et les incertitudes.

★ Méthode

Pour tester une relation linéaire entre 2 grandeurs mesurées, il faut utiliser un logiciel de régression linéaire **intégrant les barres d'incertitude** :

- Pour vérifier que les valeurs mesurées sont compatibles avec un modèle affine, il faut observer l'alignement des points sur la droite de régression.
- Pour vérifier que les incertitudes-types sont compatibles avec le modèle affine, il faut vérifier que la droite de régression passe suffisamment près des points de mesures en tenant compte des barres d'incertitudes.

Remarques

- Lors d'une régression linéaire avec présence de barres d'incertitude, Regressi donne une incertitude sur les coefficients de la droite d'ajustement a et b . Cette incertitude est une incertitude élargie correspondant à un niveau de confiance de 95%.
- On peut également utiliser une méthode numérique pour évaluer les incertitudes-types sur les coefficients a et b du modèle. Cette méthode appelée Monte-Carlo consiste à générer des ensembles de nouvelles mesures (x_i, y_i) puis réaliser une nouvelle régression linéaire. En réalisant un grand nombre de fois cette opération, on obtient les valeurs des incertitudes-types sur les coefficients a et b du modèle (en prenant les écarts-types des distributions de a et b obtenues).

Une régression complète ne se limite pas à l'obtention des paramètres de la droite de régression : il est indispensable de valider le modèle estimé, c'est-à-dire de vérifier son adéquation aux données analysées.

Méthode : analyse des résidus

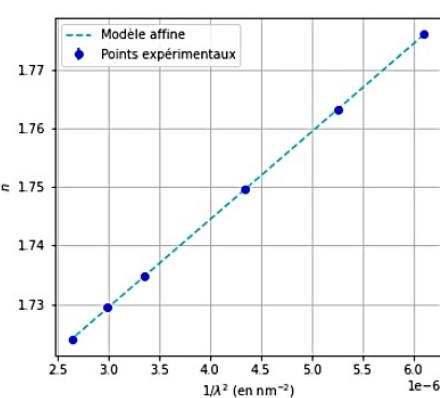
Résidus e : écarts « verticaux » entre les points de mesure et la droite de régression.

Ainsi, si la droite de régression a pour équation $y = ax + b$: à chaque valeur x_i on peut associer l'estimation $ax_i + b$ de Y donnée par la droite de régression, et la comparer à la valeur observée y_i : on obtient ainsi le résidu $e_i = y_i - (ax_i + b)$.

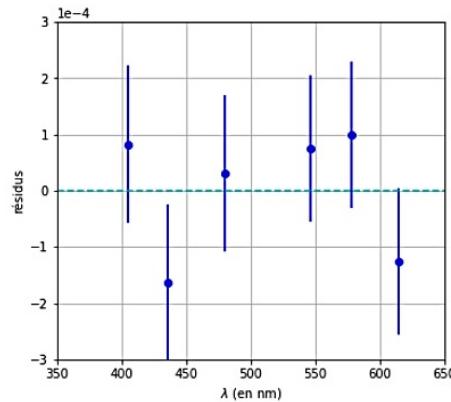
- Si le point expérimental appartient à la droite de régression (cas « idéal »), on obtient un résidu nul : $e_i = 0$;
- Sinon, on regarde si ces résidus sont bien aléatoirement distribués autour du 0 et si ils sont bien inférieurs 2 fois la valeur de l'incertitude-type sur y_i .

Remarques

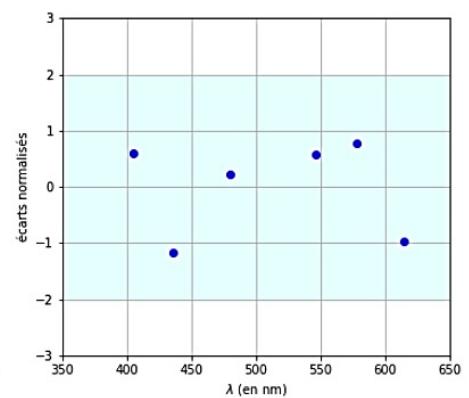
- si $e_i \gg 2u(y_i)$, on peut considérer que le point est aberrant (attention ! il faut le documenter !) et l'éliminer de la liste.
- Pour visualiser facilement ces différents cas, on représente graphiquement les résidus e_i en fonction des x_i , en mettant en évidence les barres d'erreur $u(y_i)$ → graphique ② ci-après.
- On peut aussi calculer les résidus normalisés : $\frac{e_i}{u(y_i)}$, ce qui revient à calculer les z-scores de chaque y_i que l'on souhaite comparer à la valeur estimée avec la régression linéaire. Pour valider l'adéquation d'un modèle linéaire pour le nuage de points expérimentaux, il faut que les z-scores soient bien tous inférieurs à 2 → graphique ③ ci-après.



①



②



③

IV Méthode Monte-Carlo pour estimer les incertitudes-types

IV.1 Estimation d'une incertitude-type composée par la méthode Monte-Carlo

Soit une grandeur physique Z définie par une fonction de la variable expérimentale X . La méthode de Monte-Carlo consiste à calculer un grand nombre de valeurs de Z pour différentes valeurs des X tirées aléatoirement dans l'intervalle $[x_{\text{mes}} - \Delta, x_{\text{mes}} + \Delta]$ avec Δ la précision sur x donnée par $\Delta = \sqrt{3}u(X)$ (pour une distribution uniforme) et $u(X)$ l'incertitude-type évaluée lors de la mesure de x . Le script déduit ensuite de l'ensemble des valeurs z_i obtenues la valeur moyenne et l'écart-type sur la distribution.



Remarques

- Dans le cas où la relation donnant Z fait intervenir plus d'une variable, la méthode reste la même, il faut simuler des jeux de données des différentes variables pour calculer ensuite la distribution de valeurs de Z .
- En l'absence d'informations pour savoir de quelle façon générer les valeurs des points simulés, on choisit une distribution de probabilité uniforme.

★ Principe

On cherche à estimer une grandeur y donnée par $y = f(x_1, x_2, \dots)$ avec les x_i des données résultant d'une mesure et f une fonction connue.

Chaque x_i est caractérisé par sa valeur et son incertitude-type $u(x_i)$.

① Fixer un nombre N de simulations à réaliser ;

② Pour chaque simulation k , k compris entre 1 et N :

- réaliser un tirage aléatoire pour chaque x_i en tenant compte de sa variabilité ;
- utiliser les valeurs de ce tirage et la fonction f pour calculer la valeur y_k associée au tirage ;
- sauvegarder cette valeur y_k ;

③ Déterminer la moyenne des y_k , qui donne la valeur de y , et l'écart-type de la distribution des y_k , qui donne l'incertitude-type de y .

```

1: import numpy as np
2: import matplotlib.pyplot as plt
3:
4: # Données expérimentales
5: lambd = 632.8 # nm
6: D = 1.5 # m
7: L = 46 # mm
8:
9: # Précisions
10: Deltalambd = 1 # nm
11: DeltaD = 0.01 # m
12: DeltaL = 2 # mm
13:
14: # Fonction de composition
15: def a(lambd,D,L):
16:     return 2*D*lambd/(L)
17:
18: # Nombre de simulation que vous voulez effectuer
19: N = 100000
20:
21: # Calculs avec une distribution de probabilité uniforme
22: long = []
23:
24: for i in range(0,N):
25:     tlambd = np.random.uniform(lambd-Deltalambd,lambd+Deltalambd)
26:     tD = np.random.uniform(D-DeltaD,D+DeltaD)
27:     tL = np.random.uniform(L-DeltaL,L+DeltaL)
28:     x=a(tlambd,tD,tL)
29:     long.append(a(tlambd,tD,tL))
30:
31: # Calcul et affichage moyenne et écart type
32: moy = np.mean(long)
33: std = np.std(long,ddof=1)
34: print("Moyenne = {:.2f} µm".format(moy))
35: print("Ecart type = {:.2f} µm".format(std))

```

Exemple : Script Python permettant de déterminer la largeur de la fente et l'incertitude-type associée lors de l'étude du phénomène de diffraction

IV.2 Modélisation affine avec la méthode de Monte-Carlo

Lorsqu'on souhaite évaluer les incertitudes sur les paramètres a et b d'une modélisation affine des grandeurs expérimentales x et y (telle que $y = ax + b$), on peut effectuer une simulation Monte-Carlo en réalisant un très grand nombre de régressions linéaires sur des séries de points générés aléatoirement à partir de chaque point x_i, y_i , à l'aide de leurs incertitudes-types expérimentales $u(x_i)$ et $u(y_i)$. Rappel :

pour une distribution uniforme de demi-étendue Δ : $u(X) = \frac{\Delta}{\sqrt{3}}$.

Les valeurs finales de la pente et de l'ordonnée à l'origine sont les moyennes de toutes leurs valeurs, et leurs incertitudes-types sont les écarts-types de ces deux ensembles de valeurs.

★ Principe

On souhaite évaluer les incertitudes sur les paramètres a et b d'une modélisation affine des grandeurs expérimentales x et y (telle que $y = ax+b$). On note m le nombre de points de mesures.

- ① Réaliser une régression linéaire unique pour estimer la pente et l'ordonnée à l'origine.
- ② Fixer un nombre N de simulation très grand (typiquement $N = 100000$).
- ③ Créer deux listes vides pour stocker les pentes et les ordonnées à l'origine des régressions.
- ④ Pour chaque i compris entre 1 et N , réaliser :
 - pour chaque j compris entre 1 et m , réaliser un tirage aléatoire d'une valeur de Y_j donnée par une loi de probabilité uniforme entre $y_j - \sqrt{3} u(y_j)$ et $y_j + \sqrt{3} u(y_j)$;
 - pour chaque j compris entre 1 et m , réaliser un tirage aléatoire d'une valeur de X_j donnée par une loi de probabilité uniforme entre $x_j - \sqrt{3} u(x_j)$ et $x_j + \sqrt{3} u(x_j)$;
 - réaliser une régression linéaire sur cet ensemble (X_j, Y_j) puis ajouter dans les listes la pente et l'ordonnée à l'origine de cette régression.
- ⑤ Calculer les écarts-types des deux listes des pentes et des ordonnées à l'origine pour obtenir les incertitudes-types ; tracer sur un graphique la droite obtenue avec la pente moyenne et l'ordonnée à l'origine moyenne ;
- ⑥ Superposer sur le même graphique les points de mesures en indiquant leurs incertitudes-type sous la forme de barres d'erreur.

Résumé

