Mesures et incertitudes

Sommaire

Introduction . . . 2
Incertitudes de type A . . . 3
Incertitudes de type B . . . 4
Composition des incertitudes . . . 5
Analyse d'un résultat . . . 8
Regression linéaire . . . 10

Programme officiel

Notions et contenus	Capacités exigibles
Variabilité de la mesure d'une grandeur physique. Incertitude. Incertitude-type.	Identifier les incertitudes liées, par exemple, à l'opérateur, à l'environnement, aux instruments ou à la méthode de mesure. Procéder à l'évaluation d'une incertitude-type par une approche statistique (évaluation de type A). Procéder à l'évaluation d'une incertitude-type par une autre approche que statistique (évaluation de type B). Associer un intervalle de confiance à l'écart-type dans l'hypothèse d'une distribution suivant la loi normale.
Incertitudes-types composées.	Évaluer l'incertitude-type d'une grandeur s'exprimant en fonction d'autres grandeurs, dont les incertitudes-types sont connues, à l'aide d'une somme, d'une différence, d'un produit ou d'un quotient. Comparer entre elles les différentes contributions lors de l'évaluation d'une incertitude-type composée. Capacité numérique : simuler, à l'aide d'un langage de programmation ou d'un tableur, un processus aléatoire permettant de caractériser la variabilité de la valeur d'une grandeur composée.
Écriture du résultat d'une mesure.	Écrire, avec un nombre adapté de chiffres significatifs, le résultat d'une mesure.
Comparaison de deux valeurs ; écart normalisé.	Comparer deux valeurs dont les incertitudes-types sont connues à l'aide de leur écart normalisé. Analyser les causes d'une éventuelle incompatibilité entre le résultat d'une mesure et le résultat attendu par une modélisation.
Régression linéaire.	Utiliser un logiciel de régression linéaire afin d'obtenir les valeurs des paramètres du modèle. Analyser les résultats obtenus à l'aide d'une procédure de validation : analyse graphique intégrant les barres d'incertitude ou analyse des écarts normalisés. Capacité numérique : simuler, à l'aide d'un langage de programmation ou d'un tableur, un processus aléatoire de variation des valeurs expérimentales de l'une des grandeurs – simulation Monte-Carlo – pour évaluer l'incertitude sur les paramètres du modèle.

Introduction

Quelques définitions

- **Mesure :** procédure expérimentale qui conduit à attribuer un ensemble de valeurs numériques à une grandeur, accompagné d'une unité appropriée. Le résultat d'une mesure décrit cet ensemble de valeurs en le complétant par des explications sur la manière dont elles ont été obtenues.
- **Observation ou indication** : bien qu'on ne s'en aperçoive pas toujours, si on répète une mesure, on trouve souvent une valeur numérique différente. Chacune de ces valeurs numériques est une observation ou indication.

Exemple : la mesure d'une tension à l'aide d'un voltmètre donne une valeur unique, une indication. Mais si on prend d'autres voltmètres de la même marque, on obtient d'autres indications.

Représentation d'une mesure

Lorsque l'on répète plusieurs fois une même mesure, la question de leur représentation se pose immédiatement. On peut utiliser :

- Un tableau :

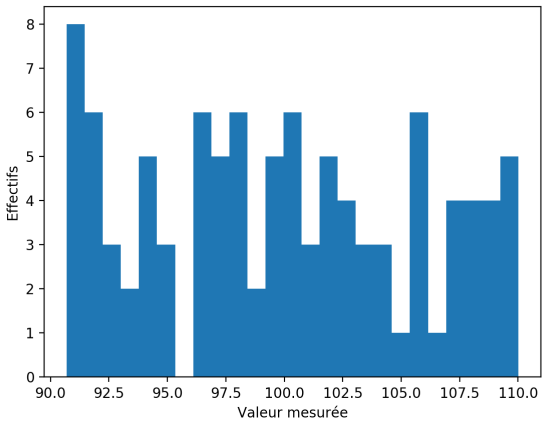
Mesure	1	2	3	4	...
Valeur	102,5	102,2	100,7	97,4	...

- Un histogramme :

```python
import numpy as np
import matplotlib.pyplot as plt

# Données stockées dans un fichier
data = np.loadtxt("...")
# ... ou écriture directe :
data = [102.5, 102.2, ... ]

# Histogramme
plt.hist(data, bins="rice")
plt.xlabel("Grandeur mesuree")
plt.ylabel("Effectifs")
plt.show()
```



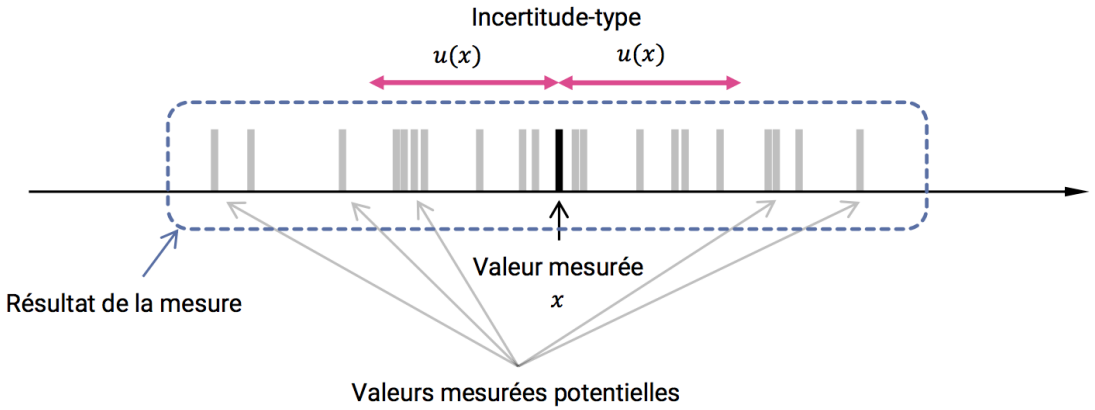
Comment écrire le résultat d'une mesure ?

Valeur mesurée et incertitude-type

Lorsque l'on ne dispose que d'une observation, la valeur mesurée X est nécessairement cette observation.

Lorsqu'on répète des observations et qu'on constate qu'elles varient, on choisit leur moyenne arithmétique comme meilleur estimateur de X

L'incertitude-type $u(X)$ est l'estimation de la dispersion des valeurs raisonnablement attribuables à une grandeur mesurée. La valeur mesurée est une de ces valeurs. L'incertitude-type quantifie donc la variabilité potentielle de cette unique valeur mesurée.



La dispersion est inhérente au processus de mesure. L'incertitude-type permet de la quantifier. Une grande incertitude ne signifie donc pas que la mesure est mauvaise !

> Les méthodes de détermination de l'incertitude type associée sont présentées supra.

Convention d'écriture

La valeur mesurée ainsi que l'estimation de l'incertitude type n'étant pas certaines, on utilise les conventions suivantes pour l'écriture des résultats pour une grandeur X :

- Afin de limiter les erreurs d'arrondis, l'incertitude type $u(\mathrm{X})$ est écrite à deux chiffres significatifs.
- Le dernier chiffre écrit de la valeur mesurée correspond à la plus petite puissance de dix sur $u(\mathrm{X})$.

Exemple :

X étant une longueur en cm, on obtient :

$$\mathrm{X} = 100.26399952677632 \text{ et } u(\mathrm{X}) = 0.29262542331105657$$

On écrira le résultat

$$\boxed{\mathrm{X} = 100,26 \text{ cm} \qquad \text{avec} \qquad u(\mathrm{X}) = 0,29 \text{ cm}}$$

On rencontre également les écritures $\mathrm{X} = 100,26 \pm 0,29$ cm ou $\mathrm{X} = 100,26(0,29)$ cm.

Idéalement on précise lors de l'écriture du résultat les méthodes utilisées pour l'obtenir (nombre d'observations, méthode de calcul de l'incertitude, ...).

Erreur ou incertitude ?

L'erreur de mesure, ou plus simplement l'erreur, est la différence entre la valeur mesurée d'une grandeur et une valeur de référence, qui peut être conventionnelle (la célérité de la lumière dans le vide) ou résulter d'une mesure dont l'incertitude est négligeable.

Elle est à distinguer de l'erreur humaine, multifactorielle : ce peut être une mauvaise saisie d'une donnée, une erreur de lecture liée à un défaut de parallaxe, l'oubli d'une grandeur d'influence dans un protocole, la non prise en compte d'un des éléments d'un montage électrique, ...

Incertitude de répétabilité (type A)

On considère le cas où l'on a effectué plusieurs observations $\{X_i\}$, et qu'elles ne sont pas toutes identiques. Une évaluation de nature statistique peut alors être envisagée.

Valeur mesurée et incertitude-type

Le meilleur estimateur de la valeur mesurée X est alors la moyenne arithmétique $\overline{X}$ des N observations X_i :

$$\boxed{\overline{X} = \frac{1}{N}\sum_{i=1}^{N} X_i}$$

Ce faisant on espère compenser partiellement les observations « trop grandes » et les « trop petites ».

L'écart type expérimental s_X, indication de la dispersion ou de l'étalement des observations est défini par

$$s_X = \sqrt{\frac{1}{n-1}\sum_{i=1}^{n}(X_i - \overline{X})^2}$$

L'incertitude type est estimée par $\boxed{u(X) = \frac{s_X}{\sqrt{N}}}$

En pratique

Avec python

```
import numpy as np
data = np.array([102.5, 102.2 , 100.7, 97.4, 99.3 , 102.2])  # Les valeurs Xi
Valeur = np.average(data) # Calcul de la moyenne
Sigma = np.std(data, ddof = 1)  # Calcul écart type
u_Valeur = Sigma/np.sqrt(len(data))       # Calcul de l'incertitude type
# ddof = 1 permet de  s' assurer que le calcul utilise bien le dénominateur N-1
print("Valeur mesurée =  ",Valeur)
print("Ecart type = ", Sigma )
print("Incertitude type = ",  u_Valeur)

>>>Valeur mesurée =   100.716666667
>>>Ecart type =  2.0292034562
>>>Incertitude type =  0.828418841998
```

Ce qui donne le résultat $X = 100,71$ avec $u(X) = 0,83$

Avec le logiciel gum_mc

Le logiciel gum_mc permet de calculer une incertitude de type A.

- Écrire dans la case « Symbole de la grandeur de sortie » le nom de la grandeur mesurée (par exemple X_{mes}) et dans la case « Expression en fonction des grandeurs d'entrée » celui des observations (par exemple x). Valider
- Cliquer sur « Ajouter une source d'erreur » puis choisir « Evaluation de type A ». Remplir les observations et valider.

Incertitude de type B

On se place ici dans le cas où on n'a pu réaliser qu'une observation unique, ou bien dans le cas où la répétition des observations conduit exactement à la même valeur. La variabilité intrinsèque de la mesure est masquée par l'appareil employé.

> Par exemple la mesure de la longueur d'une table avec une règle graduée en dm conduit toujours au même résultat.

Construction d'un intervalle de valeurs représentant une mesure

On présente quelques exemples (liste non exhaustive) de construction d'un intervalle de mesure.

- On utilise une balance avec une précision de 1 g. On lit $m = 58$ g. L'intervalle de mesure est [57 g, 59 g].
- On mesure avec une règle graduée en mm une longueur de $12,3$ cm. On peut associer un intervalle de mesure d'une demi-graduation de part et d'autre, soit $[12,25 \text{ cm}; 12,35 \text{ cm}]$.
- On réalise l'image optique sur un écran. On constate que l'image est nette sur la plage $[14,6; 16.2 \text{ cm}]$. On retient comme mesure la valeur moyenne $15,4$ cm.
- On mesure une tension à l'aide d'un voltmètre. La notice indique une précision de $\pm\,0,5\%$. On lit U = $14,3$ V. L'intervalle de mesure est $[14,3 * (1 - 0,5/100); 14,3 * (1 + 0,5/100)] = [14,2285; 14,3715]$.

Incertitude-type associée à l'intervalle de valeurs obtenu

Pour calculer l'incertitude type de l'ensemble des valeurs comprises dans l'intervalle défini, on fait le plus souvent l'hypothèse que ces valeurs y sont équiréparties. Autrement dit, si on représentait un histogramme avec un grand nombre de valeurs dans cet intervalle, celui-ci serait rectangulaire. On peut alors montrer que pour un intervalle $[\mathrm{X} - a, \mathrm{X} + a]$ l'incertitude type est

$$\boxed{u(\mathrm{X}) = \frac{a}{\sqrt{3}}}$$

Exemples

- Pour la mesure de X à l'aide d'une règle graduée au mm, $u(\mathrm{X}) = 0,5/\sqrt{3} = 0,29$ mm. Le résultat est indépendant de la valeur mesurée.
- Pour la mesure U = $14,3$ V avec un voltmètre de précision $0,5\%$, $u(\mathrm{X}) = 0,5/100 * 14,3/\sqrt{3} = 0,041$ V. Le résultat dépend de la valeur mesurée.
- Pour la mesure de la position d'un écran X_e avec un intervalle estimé de $\pm\,1$ cm, $u(\mathrm{X}_e) = 1/\sqrt{3} = 0,58$ cm. Deux expérimentateurs peuvent conclure à deux valeurs différentes de l'intervalle initial.

Avec le logiciel gum

Le résultat précédent est basé sur l'hypothèse de valeurs équiréparties. Le logiciel gum permet d'envisager d'autres distributions (gaussienne par exemple) mais les résultats ne sont pas à mémoriser.

- Écrire dans la case « Symbole de la grandeur de sortie » le nom de la grandeur mesurée (par exemple $\mathrm{X}_{\mathrm{mes}}$) et dans la case « Expression en fonction des grandeurs d'entrée » celui des observations (par exemple x). Valider
- Remplacer l'estimateur (initialement fixé à 0) par la valeur mesurée.
- Cliquer sur « Ajouter une source d'erreur » puis choisir « Evaluation de type B ». Remplir les observations et valider. Par défaut la distribution proposée est rectangulaire.

Composition des incertitudes

Dans cette fiche on présente sur un exemple les méthodes permettant de déterminer une grandeur calculée à partir d'une ou plusieurs grandeurs mesurées. Cela nécessite de composer les incertitudes types.

On mesure une vitesse c à partir d'une distance D et une durée T, en appliquant la relation

$$c = \frac{\mathrm{D}}{\mathrm{T}}$$

Évaluation par calcul direct

On dispose de N couples d'observations (D, T). On calcule les N valeurs c_i associées. On se ramène dès lors à une étude statistique de type A avec

$$\boxed{c_{\text{exp}} = \frac{1}{\mathrm{N}}\sum_{i=1}^{\mathrm{N}} c_i \qquad \text{et} \qquad u(c_{\text{exp}}) = \frac{s_{c_{\text{exp}}}}{\sqrt{\mathrm{N}}}}$$

Évaluation par simulation Monte-Carlo

On dispose de la mesure de D et de son incertitude type $u(\mathrm{D})$, ainsi que de celle de T et de $u(\mathrm{T})$. On génère numériquement un ensemble de valeurs pour c obtenues en tirant aléatoirement un grand nombre couple (D,T) dans les intervalles $[\mathrm{D} - u(\mathrm{D}); \mathrm{D} + u(\mathrm{D})]$ et $[\mathrm{T} - u(\mathrm{T}); \mathrm{T} + u(\mathrm{T})]$. On se ramène alors à un calcul de type A.

Avec python

```
# Les bibliothèques
import numpy as np
import numpy.random as rd

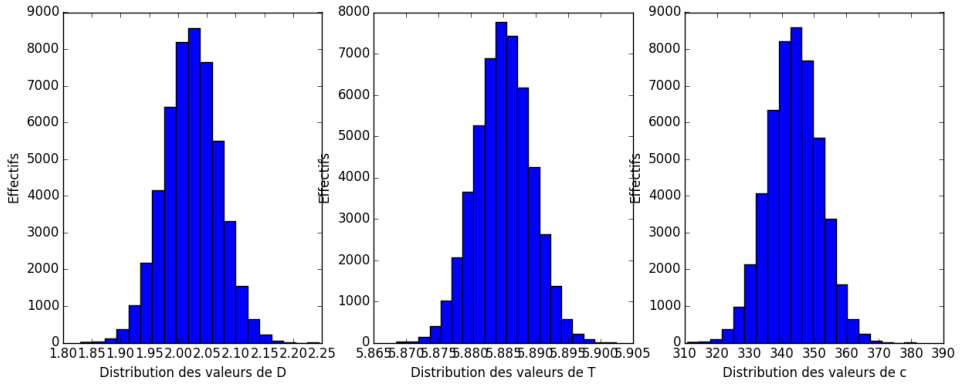
# Entrée des données
D , uD =  2.023 , 0.047 # La distance et son incertitude type (en m)
T,  uT = 5.8850E-3 , 0.0043E-3 # La durée et son incertiutde type (en s)

# Simulations : on génère deux tableaux numpy aléatoirement avec N valeurs.
N = 50000 # nombre de simulations Monte-Carlo
D_MC = rd.normal(D, uD, N)
T_MC = rd.normal(T, uT, N)

# On calcule les N vitesses associées (terme à terme, ce sont des tableaux numpy)
C_MC = D_MC/T_MC

# On calcule la moyenne et l'écart-type :
C_moy = np.average(C_MC)
uC= np.std(C_MC, ddof=1)
print("C = ",C_moy, " +/- " , uC)
```

Les distributions de valeurs de D, T et C sont représentées ci-dessous :



Évalution par l'utilisation d'une formule mathématique

On peut également utiliser des règles de propagation des incertitudes pour estimer au mieux l'incertitude d'une grandeur mesurée indirectement.

Quelques règles de calcul

Sont au programme de PCSI les règles suivantes :

- **g est une somme ou une différence**

$$\boxed{g = x \pm y \qquad \Longrightarrow \qquad u(g) = \sqrt{u(x)^2 + u(y)^2}}$$

- g est **un produit ou un quotient** :

$$\boxed{g = x\,y \quad \text{ou} \quad \frac{x}{y} \qquad \Longrightarrow \qquad \frac{u(g)}{g} = \sqrt{\left(\frac{u(x)}{x}\right)^2 + \left(\frac{u(y)}{y}\right)^2}}$$

> Le rapport $u(g)/g$ est l'incertitude type relative. Elle s'exprime usuellement en pourcentage.

Condition de validité

Les règles précédentes supposent :

- que les mesures de x et de y sont indépendantes (les grandeurs ne sont pas corrélées).
- Que les incertitudes types sont faibles. En pratique l'incertitude type relative doit être de l'ordre de quelques pourcents. Cela se justifie en étudiant la différentielle de la fonction de deux variables $g(x, y)$.

Sur l'exemple du calcul de vitesse

Avec les données fournies précédemment, on obtient $\dfrac{u(\mathrm{D})}{\mathrm{D}} = 2,3\%$ et $\dfrac{u(\mathrm{T})}{\mathrm{T}} = 0,07\%$. Le calcul de propagation donne $\sqrt{\left(\dfrac{u(\mathrm{D})}{\mathrm{D}}\right)^2 + \left(\dfrac{u(\mathrm{T})}{\mathrm{T}}\right)} = 2,3\%$! L'incertitude relative finale est quasiment égale à celle portant sur la mesure de D.

La vitesse obtenu est $c = \mathrm{D}/\mathrm{T} = 343,755310$ m/s. Son incertitude type est alors

$$u(c) = c \times \frac{u(\mathrm{D})}{\mathrm{D}} = 3,947803 \text{ m/s}$$

On écrit le résultat final en faisant attention aux chiffres significatifs :

$c = 343,8$ m/s avec $u(c) = 7,9$ m/s (propagation)

> On retient qu'avant de faire ce calcul, il convient de comparer les incertitudes relatives entre elles. On peut très souvent se limiter en pratique à la plus grande. Pour un calcul de somme ou de différence il faut comparer les incertitudes types directement.

Utilisation du logiciel Gum_mc

- Ecrire dans la case « Symbole de la grandeur de sortie » le nom de la grandeur mesurée (ici c) et dans la case « Expression en fonction des grandeurs d'entrée » celui des observations (ici D/T). Valider
- Pour chaque observable procéder à l'évaluation de l'estimateur et de l'incertitude associée selon les méthodes vues précédemment (type A ou type B).
- Cliquer sur Valider et Calculer la grandeur de sortie.
- On obtient alors les incertitudes-type par propagation ou par simulation Monte-Carlo.

Analyse d'un résultat

Disposant du résultat d'une expérience on peut être amené à :

- le comparer à une autre mesure réalisée lors d'une autre expérience. Cette situation est rencontrée lorsque l'on mesure une même grandeur par deux protocoles différents.
- le comparer à une valeur de référence qui est soit une valeur fixée (par exemple la célérité de la lumière dans le vide) soit une valeur tabulée (par exemple une valeur fournie par un constructeur).

Comparaison à une valeur de référence

La méthode de décision relève des tests statistiques. On considère dans un premier temps que la grandeur mesurée est décrite par la valeur de référence, et que le processus de mesure y ajoute une variabilité, caractérisée par l'incertitude-type. Sous cette hypothèse, la valeur mesurée est susceptible de s'éloigner de plusieurs écarts-types de la valeur de référence, mais si cela arrive c'est un évènement rare. Aussi, si on observe un éloignement considérable, on décide de rejeter l'hypothèse initialement formée : la grandeur mesurée n'est pas décrite par la valeur de référence.

> Par exemple si l'on vous propose de jouer à pile ou face en vous disant que la pièce n'est pas truquée mais qu'au bout de 10 lancers vous tombez à chaque fois sur face il est sage de conclure qu'en réalité la pièce est truquée.

On dispose d'une mesure $\mathrm{X_{mes}}$ d'incertitude type $u(\mathrm{X_{mes}})$ que l'on souhaite comparer à une valeur de référence $\mathrm{X_{ref}}$. On calcule **l'écart normalisé** (ou z-score) défini par

$$\boxed{\mathrm{Z} = \left| \frac{\mathrm{X_{mes}} - \mathrm{X_{ref}}}{u(\mathrm{X_{mes}})} \right|}$$

En TP, On considère la mesure comme validée si

$$\boxed{\mathrm{Z} < 2}$$

> La valeur 2 est un seuil raisonnable qui correspond à un « risque de se tromper » inférieur à 5%. Sa valeur reste **arbitraire** et dépend en particulier du domaine d'étude. Par exemple en physique des particules le seuil utilisé pour valider une découverte est de 5 (le risque de s'être trompé est alors beaucoup plus faible).

Comparaison entre deux mesures

On dispose d'une mesure $\mathrm{X_{mes,1}}$ d'incertitude type $u(\mathrm{X_{mes,1}})$ que l'on souhaite comparer à une seconde mesure $\mathrm{X_{mes,2}}$, d'incertitude type $u(\mathrm{X_{mes,2}})$. L'écart normalisé est alors défini par

$$\boxed{\mathrm{E} = \left| \frac{\mathrm{X_{mes,1}} - \mathrm{X_{mes,2}}}{\sqrt{u^2(\mathrm{X_{mes,1}}) + u^2(\mathrm{X_{mes,2}})}} \right|}$$

On considère que les deux mesures sont compatibles entre elles si

$$\boxed{\mathrm{E} < 2}$$

Tout comme la comparaison à une valeur de référence ce seuil est arbitraire et dépend du domaine d'étude.

Que faire en cas de désaccord ?

Il convient en cas de désaccord de se poser les questions suivantes (liste non exhaustive) :

- Les mesures ont-elles été correctement réalisées ? Par exemple une mesure d'angle nécessite de connaitre le zéro, on a pu faire une erreur de parallaxe en se positionnant mal, on a confondu rayon et diamètre...
- Les incertitudes ont-elles été correctement évaluées ? N'ont-elles pas été sous-estimées ? Une incertitude-type n'a-t-elle pas été oubliée ?
- Le protocole de mesure est-il correct ? Par exemple certains protocoles nécessitent de réaliser une correction de la valeur lue pour prendre en compte l'influence de l'appareil de mesure, ou de ré-étalonner un capteur en fonction de la température.
- N'y a-t-il pas une (ou plusieurs) valeurs aberrantes ? Auquel cas il convient de les exclure (en expliquant pourquoi, par exemple une erreur de report), ou de refaire l'expérience associée si c'est possible.
- La modélisation du phénomène proposé est-elle correcte ? Si l'on parvient à trouver une cause de désaccord alors on peut refaire l'expérience en en tenant compte. Par exemple en mécanique on peut prendre en compte des frottements que l'on aura négligé de prime abord.
- Au bout de la démarche peut-être venez vous de mettre en évidence un nouveau phénomène !

> **Une illustration :**
> En 2011 des chercheurs de l'expérience OPERA ont mesuré la vitesse des neutrinos et sont arrivés à une valeur supérieure à celle de la vitesse de la lumière dans le vide (cette nouvelle a fait la une du journal de 20 h). Mais les dits chercheurs, surpris par leur mesure, ont mis à disposition toutes leurs données afin que la communauté scientifique puisse les analyser. Il s'avère qu'un câblage défectueux modifiait le temps de parcours mesuré, suffisamment pour que le résultat soit en réalité une vitesse en accord avec la théorie de la relativité (nouvelle qui n'est pas du tout passée au journal de 20 h mais est restée cantonnée à la communauté scientifique...) .

Regression linéaire sur un exemple

Introduction

Dans de très nombreuses situations, on peut changer un paramètre d'une expérience ce qui permet de tester un ensemble plus large de valeurs. Dès lors utiliser simplement une seule expérience est réducteur.

On est également amené à vouloir tester des modèles plus ou moins élaborés : le phénomène étudié est-il modélisable par une loi de puissance ? Par une exponentielle ?

La regression linéaire est une des méthodes d'étude d'un ensemble de points expérimentaux couramment employée.

Exemple d'étude

On présente la méthode sur l'exemple de l'étude de la loi de Cauchy : dans le domaine visible, l'indice optique n d'un verre varie avec la longueur d'onde selon la loi

$$n(\lambda) = \mathrm{A} + \frac{\mathrm{B}}{\lambda^2}$$

On cherche les valeurs expérimentales des paramètres A et B d'un verre, sachant que l'on dispose des mesures suivantes :

λ (en nm)	404,7	435,8	480,0	546,1	578,1	615
n	1,7761	1,7632	1,7496	1,7348	1,7294	1,7240
u(n)	0,0014	0,0014	0,0014	0,0013	0,0013	0,0013

Les valeurs de λ sont supposées tabulées.

Déterminer les paramètres A et B

Pour déterminer A et B, on effectue une regression linéaire en cherchant pour l'ensemble des couple de mesure (x_i, y_i), le jeu de valeurs (A, B) permettant de se rapporcher le plus possible de la loi $y_i = \mathrm{A}\,x_i + \mathrm{B}$. Sur l'exemple proposé $y_i = n_i$ et $x_i = 1/{\lambda_i}^2$. On montre alors que les valeurs de A et B sont celles qui rendent minimum la grandeur

$$\sum_{i=1}^{n} \left(y_i - (\mathrm{A} + \mathrm{B}\,x_i)\right)^2$$

En exploitant la condition de minimisation, on obtient deux formules (qui ne sont pas à connaître) donnant A et B. Elles sont implémentées informatiquement :

Avec sa calculatrice

Sur les calculatrices de marque TI on utilise RégLin. Sur les Casio REG, sur les Numworks Regressions, ...

Souvent il est retourné la valeur de A et B ainsi que du coefficient de corrélation R^2 dont l'interprétation est particulièrement délicate (voir supra).

> *En physique on n'utilisera quasiment jamais cette méthode car elle ne permet pas de caractériser proprement les incertitudes ni de vérifier correctement le caractère linéaire.*

Avec regressi

Pour faire une régression linéaire, on commence par entrer les valeurs expérimentales dans le tableur. Il faut ensuite :

- **Entrer les incertitudes** : onglet **variables**, puis **incertitudes**, ou directement en cliquant sur $\boxed{u}$. Double-cliquer sur la colonne de la grandeur puis entrer la valeur ad-hoc.

  > On peut entrer des incertitudes-type différentes d'une mesure à une autre.

- **Visualisation :** menu **options**, onglet **graphique**, cocher « tracé des ellipses d'incertitudes ».
- **Modélisation :** cliquer sur l'icône **modélisation** et de choisir le modèle prédéfini. On peut imposer une régression passant par zéro (choix du modèle affine ou linéaire).
- **Prise en compte des incertitudes dans la modélisation : :** menu **options**, onglet **calculs**, cocher la case « utiliser la méthode des ellipses ». On obtient l'incertitude type sur a et b après avoir cliqué sur **Actualiser** à gauche du graphique.

Avec python, utilisation de numpy.polyfit

Deux fonctions permettent de réaliser une régression linéaire : `polyfit` de la librairie `numpy` (qui permet également des modélisation polynomiales), `linregress` de la librairie `scipy.stats`. On présente un exemple de mise en oeuvre.

```
## Importation des bibliothèques utiles
import numpy as np                    # pour le traitement vectoriel des données
import matplotlib.pyplot as plt     # pour les tracés

## Saisie des valeurs expérimentales
lamb = np.array([404.7, 435.8, 480.0, 546.1, 578.1, 615]) #longueurs d'onde (en nm)
n = np.array([1.77610, 1.76325, 1.74957, 1.73481, 1.72942, 1.72397]) # indices n
u_n = np.array([1.4, 1.4, 1.4, 1.3, 1.3, 1.3])*1e-3    # incertitudes-type sur n

## Ajustement affine (polynome de degré 1)
Y , X = n , 1/lamb**2
p = np.polyfit(X, Y, 1)    # p est un tableau  des coefficients de l'ajustement
                           # p[0] : pente ; p[1] : ordonnée à l'origine
Yreg = p[0]*X + p[1]       # calcul des ordonnées des points "modèle"

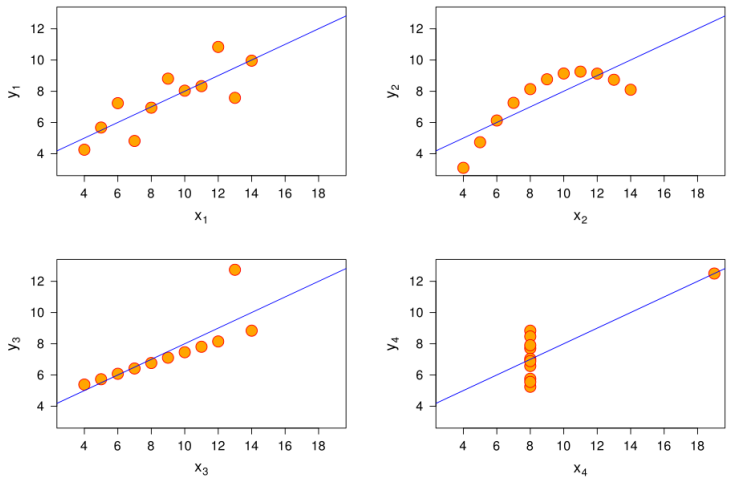
## Affichage du résultat:
print("Pente = ", p[0])
print("Ordonnée à l'origine = ", p[1])
```

Accord expérimental ?

Le problème du coefficient de corrélation

Une calculatrice fournit un outil statistique, le coefficient de corrélation r, compris entre -1 (corrélation parfaite de pente négative) et 1 (corrélation parfaite de pente positive). Si $r = 0$ il n'y a aucune corrélation. Souvent les logiciels indiquent r^2, qui doit être le plus proche de 1.

Sur l'exemple ci-dessous appelé quartet d'Anscombe, on voit quatre distributions de mesures ayant **exactement** les mêmes propriétés statistiques, en particulier le même coefficient r !



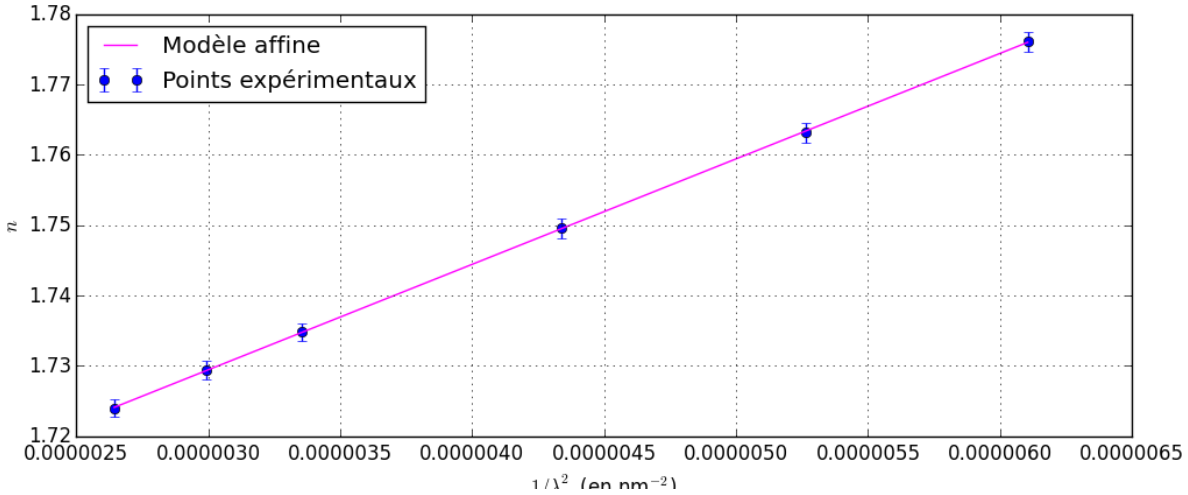
Quartet d'Anscombe : les quatre distributions ont les mêmes propriétés statistiques...

Une version encore plus visuelle est à regarder ici : https://www.youtube.com/watch?v=It4UA75z_KQ

Contrôle visuel direct

Rien ne remplace dès lors un contrôle visuel des données. Il consiste à tracer la droite de regression et à regarder si elle « passe » bien par les différents points (représentés avec leur incertitude).

```
## Tracé
plt.errorbar(X, Y, yerr = u_n, fmt = 'bo',label = "Points expérimentaux")
plt.plot(X, Yreg, 'c--',label = "Modèle affine")
plt.xlabel(r"$1/\lambda^2$ (en nm$^{-2}$)")
plt.ylabel(r"$n$")
plt.grid(), plt.legend(loc = 'best')
plt.show()
```



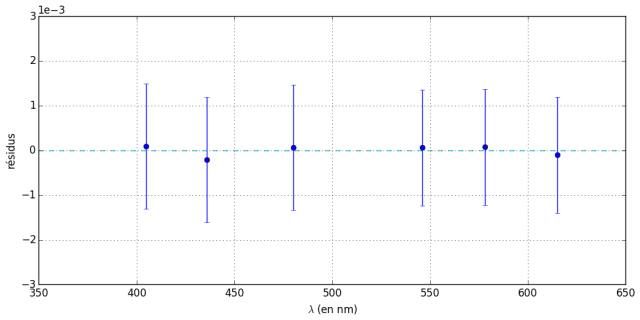
Ici la régression linéaire est validée mais l'observation directe peut être délicate lorsque les incertitudes sont faibles.

Contrôle à l'aide des résidus ou de l'écart normalisé

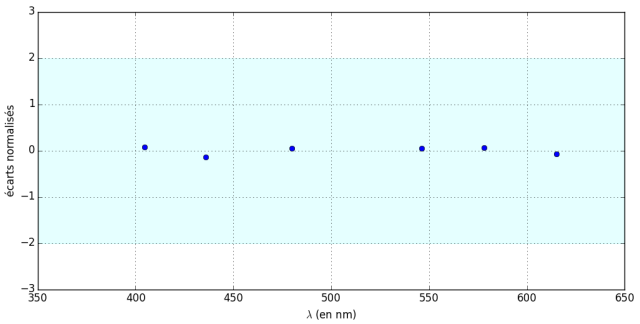
On peut aussi pour chaque mesure calculer le résidu $y_i - (\mathrm{A} + \mathrm{B}\,x_i)$, autrement dit l'écart entre la valeur attendue et la valeur mesurée. La régression est validée si les points expérimentaux sont compatibles avec un résidu nul.

Enfin on peut représenter pour chaque mesure l'écart normalisé. La regression est validée si pour chaque mesure il est inférieur à 2.

On obtient les deux figures suivantes, qui chacune permetttent de valider la regression.



Contrôle par les résidus



Contrôle par l'écart normalisé

Incertitude sur les coefficients A et B par simulation Monte-Carlo

Les données initiales étant obtenues avec une incertitude type, les valeurs de A et B ne peuvent être certaines. Pour déterminer les incertitudes sur A et B, on réalise une simulation Monte-Carlo en générant un grand nombre de jeux valeurs à partir des données et de leurs incertitudes (méthode déjà utilisée pour la composition des incertitudes). On obtient alors une distribution de valeurs de A et B dont on extrait la moyenne et l'incertitude-type.

```
Nmc=  10000 # Nombre de simulation
Amc , Bmc = [] , [] # Initialisation des listes contenant les valeurs de A et B

for i in range(NMC):    # boucle sur le nombre de simulations
  # on simule la mesure des n pour les 6 longueurs d'onde
  Yn  = Y + rd.normal(0, u_n, size = 6)
  # on reprend l'ajustement affine avec cette série de valeurs de n
  p = np.polyfit(X, Yn, 1)
  # on stocke les valeurs des paramètres d'ajustement dans les listes Amc et Bmc
  Amc.append(p[1])
  Bmc.append(p[0])

## Analyse statistique de lénsemble des valeurs de A et de B obtenues
Amoy = np.mean(AMC)
u_A = np.std(AMC, ddof = 1)
print("A = ", Amoy , " +/- ", u_A)
Bmoy = np.mean(BMC)
u_B = np.std(BMC, ddof = 1)
print("B = ", Bmoy , " +/- ", u_B)
```

L'avantage de cette méthode est qu'elle fournit à la fois les valeurs de A et B (qui sont affinées par rapport à la régression initiale) et leurs incertitudes. L'inconvenient est qu'elle est plus longue à mettre en oeuvre.

> En pratique cocher la case « Utiliser les ellipses d'incertitudes » lors d'une regression à l'aidre du logiciel regressi aboutit quasiment au même résultat que celui obtenu par simulation Monte-Carlo...

Parfois on peut éviter la regression linéaire...

Lorsque l'on cherche à modéliser un phénomène par une loi de type $y = \mathrm{A}\,x$ la regression linéaire peut être évitée. Il suffit en effet pour les N couples de valeurs mesurées (x_i, y_i) de calculer les N valeurs associées A_i puis d'effectuer un traitement statistique (type A).

On peut conserver la validation graphique par la méthode des résidus ou par la méthode de l'écart normalisé.