

Mesures et incertitudes

1 Variabilité des grandeurs physiques

La mesure d'une grandeur physique X quelconque est par nature un processus comportant une variabilité intrinsèque. Cette variabilité peut avoir de multiples causes, dont on peut donner plusieurs exemples (de manière non exhaustive).

Causes liées à l'observateur (ou expérimentateur) :

- ✓ mesure de la période d'oscillation d'un oscillateur mécanique à l'aide d'un chronomètre (détermination du passage par la verticale),
- ✓ positionnement d'un écran pour la mise au point de la netteté d'une image sur banc d'optique (critère de netteté),
- ✓ positionnement d'une lunette de visée pour le pointée d'une raie spectrale sur goniomètre,

Causes liées au système physique :

- ✓ mesure de la pression ou de la température d'un gaz (grandeurs fluctuantes d'origines microscopiques),
- ✓ mesure de la tension aux bornes d'un dipôle (existence d'une partie variable aléatoire d'origine thermique).

Causes liées à la méthode de mesure :

- ✓ mesure de la taille d'une personne à la toise (position bien verticale, mesure le matin ou le soir),
- ✓ mesure de la longueur d'une table (en quels points prendre la mesure si les bords ne sont pas parallèles ? Cette longueur est-elle unique à l'échelle atomique ?)

Cause liée à l'instrument de mesure :

- ✓ exemple de plusieurs voltmètres de même type branchés en parallèle aux bornes d'une alimentation :



On peut donc considérer que :

Toute mesure d'une grandeur physique X doit être vue comme un processus aléatoire qui conduit à la détermination d'un ensemble de valeurs $(x_0, x_1, \dots)$ associées à la grandeur X .

On peut, dans le processus de mesure, considérer X comme une variable aléatoire dont on cherche à déterminer au mieux certaines des caractéristiques.

2 Moyenne et incertitude type

2.1 Loi de probabilité continue

Les grandeurs physiques variant le plus souvent de manière continue (i.e. elles peuvent prendre n'importe quelle valeur dans un intervalle), et sont caractérisées par une loi de probabilité continue.

Pour une variable aléatoire X à valeurs continues, on note dP la probabilité que la variable X prenne une valeur dans l'intervalle $[x, x + dx]$. La densité de probabilité associée à X est la fonction $\varphi(x)$ telle que :

$$dP = \varphi(x)dx$$

On a pour $\varphi(x)$ les propriétés suivantes :

$$\int_{\mathbb{R}} \varphi(x)dx = 1$$

$$P(X \leq a) = \int_{-\infty}^a \varphi(x)dx$$

Parmi les différentes lois de probabilité, on peut citer :

- ✓ La loi rectangulaire (ou uniforme) ($\mu \in \mathbb{R}, \Delta \in \mathbb{R}^+$) :

$$\varphi(x) = \frac{1}{2\Delta} \text{ pour } x \in [\mu - \Delta, \mu + \Delta], \quad \varphi(x) = 0 \text{ sinon}$$

- ✓ La loi normale ($\mu \in \mathbb{R}, \sigma \in \mathbb{R}^+$) :

$$\varphi(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-(x-\mu)^2/(2\sigma^2)}$$

La loi de probabilité associée à une grandeur physique n'est pas forcément connue à l'avance, et on peut chercher à la déterminer expérimentalement. Cependant, on se limite dans la plupart des cas à la détermination de deux caractéristiques, la moyenne $\bar{X}$ et l'incertitude type (ou écart type) $u(X)$.

2.2 Moyenne

On définit la moyenne (ou espérance mathématique) $\bar{X}$ (ou $E(X)$) de la grandeur X par :

$$\bar{X} = E(X) = \int x\varphi(x)dx$$

En appliquant cette définition aux deux lois de probabilité précédentes, on obtient :

✓ $\bar{X} = \mu$ pour la loi rectangulaire

✓ $\bar{X} = \mu$ pour la loi normale

2.3 Incertitude type

L'incertitude type (ou écart type) $u(X)$ (ou σ_X) associée à X est par définition la racine de la variance, soit

$$u(X) = \sigma_X = \sqrt{E((X - \bar{X})^2)}$$

On peut en donner une formulation équivalente :

$$u(X) = \sigma_X = \sqrt{E(X^2) - \bar{X}^2}$$

L'incertitude type mesure la dispersion des valeurs que peut prendre X autour de la valeur moyenne.

Exemple de calcul pour la loi rectangulaire (ou uniforme) :

$$E((X - \bar{X})^2) = \int (x - \mu)^2 \varphi(x) dx = \frac{1}{2\Delta} \int_{\mu-\Delta}^{\mu+\Delta} (x - \mu)^2 dx$$

On pose $t = x - \mu$, d'où

$$E((X - \bar{X})^2) = \frac{1}{2\Delta} \int_{-\Delta}^{\Delta} t^2 dt = \frac{1}{6\Delta} [t^3]_{-\Delta}^{\Delta} = \frac{\Delta^2}{3}$$

On a donc pour cette distribution

$$\sigma_X = u(X) = \frac{\Delta}{\sqrt{3}}$$

Résultat admis pour la loi normale :

$$u(X) = \sigma$$

Lors d'une mesure d'une grandeur physique X , la détermination de $\bar{X}$ et de $u(X)$ permet de caractériser partiellement la mesure. La loi de probabilité de X n'étant en général pas connue, et on est obligé d'estimer ces grandeurs, soit à partir de mesures expérimentales (estimation de type A), soit à partir de modèles ou d'informations concernant l'appareil de mesure (estimation de type B).

2.4 Estimation de type A

On se place dans le cas où on a pu réaliser N mesures successives d'une grandeur X , et ainsi obtenir un ensemble de N valeurs $(x_0, x_1, \dots, x_{N-1})$ associées à la même grandeur X .

On peut alors créer un estimateur de la moyenne $\bar{X}$, noté X^* :

$$X^* = \frac{1}{N} \sum_{i=0}^{N-1} x_i$$

On peut aussi créer un estimateur de l'incertitude type

$$s_{exp} = \sqrt{\frac{1}{N-1} \sum_{i=0}^{N-1} (x_i - X^*)^2}$$

Ces grandeurs permettent d'approcher les grandeurs recherchées, à savoir $\bar{X}$ et $u(X)$, qui elles ne sont jamais accessibles (leur détermination nécessiterait de posséder une infinité de valeurs de x_i).

En pratique, on utilisera ces estimateurs pour évaluer les grandeurs $\bar{X}$ et $u(X)$, et on écrira

$$\bar{X} \approx \frac{1}{N} \sum_{i=0}^{N-1} x_i = X^*$$

$$u(X) \approx \sqrt{\frac{1}{N-1} \sum_{i=0}^{N-1} (x_i - X^*)^2} = s_{exp}$$

L'utilisation de ces estimateurs est d'autant meilleure que N est grand. Pour s'en convaincre, on peut réaliser la simulation informatique suivante en considérant une grandeur X dont les valeurs sont distribuées selon une loi uniforme $\mu = 1, \Delta = 0.1$. En réalisant un tirage aléatoire de N mesures, on peut calculer X^* et s_{exp}^2 et comparer aux valeurs $\bar{X} = \mu = 1$ et $\sigma_X^2 = \frac{\Delta^2}{3} = \frac{0.01}{3} \approx 0.003333$.

	Estimation 1, $N = 10$	Estimation 2, $N = 10^3$	Estimation 3, $N = 10^5$	Moyenne de $p = 100$ tirages, $N = 10^5$
X^*	1.030116	1.003968	1.000251	1.000006
s_{exp}^2	0.003693	0.003412	0.003334	0.003333(8)

On donne ci-dessous le code python associé aux premières simulations

```
N = 1000
mu = 1
Delta = 0.1

# tirage de N valeurs selon une loi uniforme
X = np.random.uniform(low=mu-Delta,high=mu+Delta,size=N)

#calcul des estimateurs X_etoile et s_exp_carré
X_etoile = np.mean(X)
s_exp_carré = np.std(X,ddof=1)**2 # ddof=1 permet de diviser par N-1

print("X_etoile = ",X_etoile)
print("s_exp_carré = ", s_exp_carré)
```

On donne aussi le code associé à la moyenne des p simulations de N tirages chacune :

```
N = 100000
mu = 1
Delta = 0.1
p = 100
S_moy = 0
S_incertitude_type = 0

for k in range(p):
    X = np.random.uniform(low=mu-Delta,high=mu+Delta,size=N)
    S_moy += np.mean(X)
    S_incertitude_type += np.std(X,ddof=1)**2
print("moyenne des valeurs moyennes estimées = ",S_moy/p)
print("moyenne des incertitude type au carré estimées = ",S_incertitude_type/p)
```

Incertitude type de la moyenne X^*

Lorsqu'on réalise N mesures indépendantes de la même grandeur X , la grandeur X^* associée (estimateur de la moyenne) est une variable d'incertitude type plus faible que X . En effet, on a

$$X^* = \frac{1}{N} \sum_{i=0}^{N-1} X_i$$

d'où

$$u(X^*)^2 = \frac{1}{N^2} \sum_{i=0}^{N-1} u(X_i)^2 = \frac{1}{N^2} \sum_{i=0}^{N-1} u(X)^2 = \frac{1}{N} u(X)^2$$

Ainsi, on a

$$u(X^*) = \frac{1}{\sqrt{N}} u(X) \approx \frac{1}{\sqrt{N}} s_{exp}$$

Ce résultat est assez intuitif car on peut s'attendre pour X^* à une certaine compensation des écarts à la valeur moyenne dans le calcul de la somme, et donc une dispersion moindre des valeurs prises par X^* (par rapport à celles prises par X). Si l'on peut réaliser une estimation de type A pour la mesure de X , alors on utilise X^* et $u(X^*)$ pour caractériser la mesure de X .

Remarque 1

Il pourrait être tentant pour diminuer $u(X^*)$ de prendre un nombre N de valeurs le plus grand possible, mais cette démarche est illusoire car :

- ✓ $u(X^*)$ décroît en $1/\sqrt{N}$, donc assez lentement avec N ,
- ✓ plus le nombre de mesures est important, plus les fluctuations du système sont susceptibles d'augmenter, et finalement, la mesure sera affectée par de nouvelles causes d'incertitudes.

Remarque 2

Pour la valeur moyenne estimée, on a $u(X^*) = \frac{1}{\sqrt{N}} s_{exp}$, mais pour toute nouvelle mesure de X , on aura une distribution caractérisée par $u(X)$.

2.5 Estimation de type B

Dans de nombreux cas, on n'a pas la possibilité de réaliser un échantillon de mesures suffisamment important, soit par impossibilité (la mesure affecte le système), soit parce que le système évolue trop rapidement entre deux mesures, soit tout simplement par

manque de temps. On doit alors utiliser une autre méthode pour déterminer l'incertitude de mesure, basée sur la nature de l'appareil de mesure utilisé. On peut distinguer trois cas typiques :

- ✓ Les spécifications de l'appareil de mesure précisent l'incertitude-type associée à la valeur mesurée ; c'est hélas rarement le cas
- ✓ Le fabricant indique une « précision » ou une « tolérance », termes vagues mais permettant de déterminer un intervalle contenant le résultat de la mesure
- ✓ La nature de l'appareil de mesure fait que l'on peut affirmer que la mesure appartient à un intervalle donné.

La loi de probabilité de la grandeur mesurée n'est pas connue, cependant on fait souvent l'hypothèse que la répartition est uniforme à l'intérieur de l'intervalle considéré. D'après les résultats précédents, on a donc, pour un intervalle de type $[\mu - \Delta, \mu + \Delta]$, une incertitude-type $u(X) = \Delta/\sqrt{3}$. Ce choix peut être discuté mais influe en réalité assez peu sur l'ordre de grandeur de $u(X)$.

Exemple 1 : mesure de la position angulaire d'une lunette à l'aide d'un vernier

Lors de la mesure la position angulaire d'une lunette au vernier, il est fréquent de ne pas pouvoir déterminer celle-ci mieux qu'à trois ou quatre graduations près du vernier. On a ainsi $\Delta = 2' = 2/60^\circ$. On en déduit $u(X) = \Delta/\sqrt{3} = 1/(30\sqrt{3})^\circ \approx 1.9 \times 10^{-2}^\circ$.

Exemple 2 : mesure à l'aide d'un appareil numérique

Voici un extrait de la documentation concernant la mesure d'une tension continue à l'aide du multimètre VC130-1

Range VC130-1/150-1	Précision	Resolution
200 mV	$\pm(0.5\% + 8)$	0.1 mV
2000 mV		1 mV
20 V		0.01 V
200 V		0.1 V
250 V	$\pm(0.8\% + 8)$	1 V

La colonne « précision » indique un pourcentage de la valeur lue, à laquelle il faut rajouter un nombre de fois la résolution (plus petite valeur mesurable sur le calibre). La valeur obtenue peut ensuite être interprétée comme la demi étendue de l'intervalle de mesure.

Pour une valeur lue $U = 10.00$ V dans le calibre 20 V, on a donc une demi-étendue de $\Delta = 0.005 \times 10.00 + 8 \times 0.01 = 0.13$ V. On en déduit $u(U) = 0.13/\sqrt{3} \approx 0.075$ V.

Pour une valeur lue $U = 10.0$ V dans le calibre 200 V, on obtient de même $\Delta = 0.005 \times 10.0 + 8 \times 0.1 = 0.85$ V, soit déduit $u(U) = 0.13/\sqrt{3} \approx 0.85$ V.

3 Ecriture d'un résultat de mesure

3.1 Valeur mesurée et incertitude type

Une mesure étant caractérisée par sa valeur moyenne et son incertitude type, on doit faire figurer les deux informations dans l'écriture du résultat de la mesure.

La méthode la plus claire consiste à séparer les deux informations, sous la forme :

$$\begin{array}{l} X = a \\ u(X) = b \end{array}$$

Où a et b sont les valeurs numériques obtenues lors de la mesure.

On veillera à écrire a et b en respectant la règle suivante :

- ✓ b s'écrit avec deux chiffres significatifs, afin de limiter les erreurs dues aux arrondis sur l'incertitude type,
- ✓ a s'écrit avec un nombre de chiffres significatifs compatible avec la valeur de $u(x)$, c'est-à-dire que les deux derniers chiffres significatifs de a doivent avoir la même puissance de 10 que ceux de b .

Exemple :

Pour une mesure d'indice optique, on trouve $n = 1,52345$ avec $u(n) = 2,53 \times 10^{-2}$. On écrira donc le résultat sous la forme :

$$\begin{array}{l} n = 1,523 \\ u(n) = 2,5 \times 10^{-2} \end{array}$$

On rencontre parfois une écriture plus compacte, mais plus ambiguë, du type :

$$X = a \pm b$$

La nature de valeur b est alors plus vague : s'agit-il de l'incertitude type $u(X)$, ou bien de la demi largeur de l'intervalle dans lequel X appartient (cas d'une loi uniforme), ou bien d'une autre grandeur encore, en lien avec un intervalle de confiance (notion développée ultérieurement) ? En l'absence de plus d'indications, on considèrera par défaut que dans la notation $X = a \pm b$, b désigne l'incertitude type $u(X)$.

3.2 Intervalle de confiance

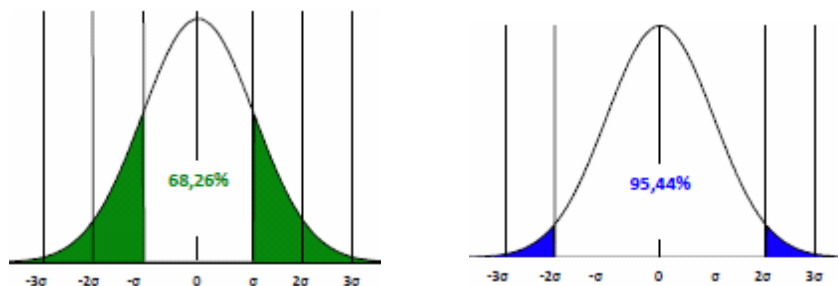
L'incertitude type $u(X)$ donne une information sur la dispersion des mesures autour de la valeur moyenne $\bar{X}$, mais l'interprétation statistique associée à $u(X)$ dépend de la nature de la loi de probabilité qui régit la variable aléatoire X . La loi normale joue ici un rôle important, car on peut montrer qu'une grandeur dépendant d'un grand nombre de variables aléatoires (ce qui est souvent le cas en pratique) suit une loi de probabilité qui est d'autant plus proche de la loi normale que le nombre de variables est grand.

On s'intéresse donc à la loi normale précédemment définie :

$$\varphi(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-(x-\mu)^2/(2\sigma^2)}$$

Avec cette loi, la probabilité P_1 que la mesure de X soit comprise dans l'intervalle $[\mu - \sigma, \mu + \sigma]$ est $P_1 = \int_{\mu-\sigma}^{\mu+\sigma} \varphi(x) dx$, et vaut $P_1 \approx 0,68$. On dit que l'intervalle $[\mu - \sigma, \mu + \sigma]$ est un intervalle de confiance à 68% pour la mesure de X . De même, la probabilité P_2 que la mesure de X soit comprise dans l'intervalle $[\mu - 2\sigma, \mu + 2\sigma]$ est $P_2 = \int_{\mu-2\sigma}^{\mu+2\sigma} \varphi(x) dx$, et vaut $P_2 \approx 0,95$. On dit que l'intervalle $[\mu - 2\sigma, \mu + 2\sigma]$ est un intervalle de confiance à 95% pour la mesure de X .

On peut représenter graphiquement ce résultat :



Remarque 1

Ceci peut aussi figurer dans l'écriture du résultat d'une mesure. Ainsi, en supposant que X suit une loi normale :

- ✓ $X = a \pm b$ à 68% signifie que $b = \sigma = u(X)$
- ✓ $X = a \pm b$ à 95% signifie que $b = 2\sigma = 2u(X)$

On retrouve le fait que l'écriture $X = a \pm b$ peut être source d'ambiguïté.

Remarque 2

Ces valeurs d'intervalle de confiance n'ont de sens que si la distribution suit une loi normale, ce qui n'est pas toujours le cas. Par exemple, pour une loi uniforme de largeur 2Δ , on a $\sigma_X = u(X) = \Delta/\sqrt{3}$ et $P_1 = \int_{\mu-\sigma_X}^{\mu+\sigma_X} \frac{1}{2\Delta} dx = 1/\sqrt{3} \approx 0,58$.

4 Propagation des incertitudes

On cherche souvent à évaluer une grandeur physique Y qui dépend d'autres grandeurs $(X_1, X_2, \dots, X_n)$, par une relation du type :

$$Y = f(X_1, X_2, \dots, X_n)$$

Pour chaque grandeur X_i , on suppose connue la valeur moyenne $\bar{X}_i$ et l'incertitude type $u(X_i)$ et on cherche à déterminer $\bar{Y}$ et $u(Y)$. On se limite ici aux cas où les grandeurs $\{X_i\}$ sont statistiquement indépendantes.

4.1 Approche mathématique

4.1.1 Formules générales

Dans le cas où les incertitudes types $u(X_i)$ sont suffisamment petites (ce qui est vérifié en général, car on cherche justement à limiter les incertitudes lors des mesures) ; on peut montrer que l'on a en première approximation :

$$\bar{Y} = f(\bar{X}_1, \bar{X}_2, \dots, \bar{X}_n)$$

$$u(Y) = \sqrt{\left(\frac{\partial f}{\partial X_1}\right)^2 u(X_1)^2 + \left(\frac{\partial f}{\partial X_2}\right)^2 u(X_2)^2 + \dots + \left(\frac{\partial f}{\partial X_n}\right)^2 u(X_n)^2}$$

On peut également donner une formule très utilisée dans le cas où Y s'exprime comme un produit des X_i à la puissance α_i , et qui permet d'exprimer l'incertitude type relative $u(Y)/|\bar{Y}|$. Pour

$$Y = X_1^{\alpha_1} X_2^{\alpha_2} \dots X_n^{\alpha_n}$$

on a après calculs

$$\frac{u(Y)}{|\bar{Y}|} = \sqrt{\alpha_1^2 \left(\frac{u(X_1)}{\bar{X}_1}\right)^2 + \alpha_2^2 \left(\frac{u(X_2)}{\bar{X}_2}\right)^2 + \dots + \alpha_n^2 \left(\frac{u(X_n)}{\bar{X}_n}\right)^2}$$

4.1.2 Exemples usuels

Différence de deux grandeurs

On s'intéresse au cas

$$Y = f(X_1, X_2) = X_2 - X_1$$

qui correspond par exemple à une mesure de distance ($d = x_2 - x_1$) ou de durée ($t = t_2 - t_1$). On a alors par application de la formule générale

$$u(Y) = \sqrt{u(X_1)^2 + u(X_2)^2}$$

Si on suppose en plus que $u(X_1) = u(X_2) = u(X)$, alors $u(Y) = \sqrt{2} u(X)$.

Par exemple, pour une mesure sur banc d'optique, on cherche à mesurer $d = x_2 - x_1$ avec $x_2 = 27,5$ cm, $x_1 = 12,4$ cm, $u(x_2) = 5$ mm et $u(x_1) = 2$ mm. On a $u(d) = \sqrt{5^2 + 2^2} = 5,4$ mm, d'où le résultat :

$$d = 15,10 \text{ cm}, \quad u(d) = 5,4 \text{ mm}$$

Rapport de deux grandeurs

On s'intéresse au cas

$$Y = f(X_1, X_2) = X_2/X_1$$

comme par exemple pour $R = U/I$.

On a pour l'incertitude type élargie :

$$\frac{u(Y)}{|\bar{Y}|} = \sqrt{\left(\frac{u(X_1)}{\bar{X}_1}\right)^2 + \left(\frac{u(X_2)}{\bar{X}_2}\right)^2}$$

soit pour l'exemple choisi :

$$\frac{u(R)}{|\bar{R}|} = \sqrt{\left(\frac{u(U)}{\bar{U}}\right)^2 + \left(\frac{u(I)}{\bar{I}}\right)^2}$$

4.1.3 Expression approchée de $u(Y)$

Il est possible de simplifier la formule de propagation avant calcul numérique, quand l'une des causes d'incertitude domine sur les autres. Prenons l'exemple d'un dosage d'une monobase par un monoacide. On a à l'équivalence :

$$c_b V_b = c_a V_{a,eq}$$

soit :

$$c_b = \frac{c_a V_{a,eq}}{V_b}$$

On a donc

$$\frac{u(c_b)}{c_b} = \sqrt{\left(\frac{u(c_b)}{c_b}\right)^2 + \left(\frac{u(V_b)}{V_b}\right)^2 + \left(\frac{u(V_{a,eq})}{V_{a,eq}}\right)^2}$$

Pour une situation typique, c_b et V_b sont mesurés à 0.5 % près environ, soit $\frac{u(c_b)}{c_b} \approx \frac{u(V_b)}{V_b} \approx 5 \times 10^{-3}$, alors que $\frac{u(V_{a,eq})}{V_{a,eq}} \approx \frac{1}{15} \approx 6,7 \times 10^{-2}$. On a donc $\left(\frac{u(c_b)}{c_b}\right)^2 \approx \left(\frac{u(V_b)}{V_b}\right)^2 \approx 2,5 \times 10^{-5} \ll \left(\frac{u(V_{a,eq})}{V_{a,eq}}\right)^2 \approx 4,4 \times 10^{-3}$. On a alors en négligeant les deux premiers termes :

$$\frac{u(c_b)}{c_b} \approx \frac{u(V_{a,eq})}{V_{a,eq}}$$

4.2 Simulation numérique de Monte Carlo

Il est parfois difficile d'appliquer les formules de propagation des incertitudes, ou bien elles ne sont pas applicables car les incertitudes type des différents X_i ne sont pas suffisamment petites. On procède alors dans ce cas à une simulation informatique de type Monte-Carlo, faisant intervenir des tirages aléatoires de valeurs.

On désire déterminer, pour une grandeur Y s'exprimant en fonction de n variables $(X_1, X_2, \dots, X_n)$ selon $Y = f(X_1, X_2, \dots, X_n)$, la valeur moyenne $\bar{Y}$ et l'incertitude type $u(Y)$. On suppose connue, pour chaque X_i , les grandeurs $\bar{X}_i$ et $u(X_i)$. Pour ce faire :

- ✓ On réalise $N \gg 1$ tirages aléatoires de chaque grandeur X_i , en utilisant une loi de probabilité caractérisée par $\bar{X}_i$ et $u(X_i)$. On obtient ainsi un ensemble de valeurs $x_{i,k}$, avec $k \in [0, N-1]$.
- ✓ On calcule les valeurs $y_k = f(x_{1,k}, x_{2,k}, \dots, x_{n,k})$ pour $k \in [0, N-1]$
- ✓ On détermine $\bar{Y}$ et $u(Y)$ à l'aide des estimateurs respectifs $y^* = \frac{1}{N} \sum_{k=0}^{N-1} y_k \approx \bar{Y}$ et $s_{exp} = \sqrt{\frac{1}{N-1} \sum_{k=0}^{N-1} (y_k - y^*)^2} \approx u(Y)$

Le choix de la loi de probabilité caractérisant chacune des variables X_i dépend de la connaissance que l'on a de cette loi. En l'absence d'information, on pourra prendre une loi uniforme.

Exemple de la différence de deux grandeurs :

Reprenons l'exemple précédent, avec $d = x_2 - x_1$ avec $x_2 = 27,5$ cm, $x_1 = 12,4$ cm, $u(x_2) = 5$ mm et $u(x_1) = 2$ mm.

On peut réaliser la simulation de Monte Carlo en utilisant le code Python suivant :

```
N = 100000
# valeurs moyennes et incertitude type de x1 et x2
x1 = 12.4
ux1 = 0.2
x2 = 27.5
ux2 = 0.5

# calcul du demi intervalle pour utilisation dans la loi uniforme
Deltax1 = ux1*3**0.5
Deltax2 = ux2*3**0.5

# tirage aléatoire de N valeurs pour x1 et x2, selon la loi uniforme
X1 = np.random.uniform(low=x1-Deltax1,high=x1+Deltax1,size=N)
X2 = np.random.uniform(low=x2-Deltax2,high=x2+Deltax2,size=N)

#Calcul de l'ensemble des valeurs de d
D = X2-X1
```

```
#Calcul de la moyenne et de l'écart type des valeurs issus des N tirages
moy = np.mean(D)
sigma = np.std(D,ddof=1)

print('resultats monteCarlo')
print('Moyenne =', moy)
print('Ecart type =', sigma)
```

Les résultats de cette simulation dépendent bien sûr un peu des tirages effectués, mais pour $N \gg 1$, ils varient assez peu et on obtient :

```
resultats monteCarlo
Moyenne = 15.098119282936814
Ecart type = 0.5372080723163197
```

On obtient donc $\bar{d} = 15,10$ cm et $u(d) = 5,4$ mm. Les résultats sont identiques à ceux obtenus par la formule mathématique (ce qui est normal car ici $|u(X_i)| \ll |X_i|$).

Exemple de la mesure de la valeur en eau d'un calorimètre

Une mesure classique de calorimétrie en TP consiste à placer une masse m_1 d'eau à la température T_1 dans un calorimètre, de rajouter une masse m_2 d'eau à la température T_2 , et d'homogénéiser et d'attendre l'équilibre thermique, de température T_f . La valeur en eau du calorimètre $\mu = C_{calo}/c_{eau}$ est alors donnée par :

$$\mu = \frac{m_1(T_1 - T_f) + m_2(T_2 - T_f)}{T_f - T_1}$$

On mesure m_1, m_2, T_1, T_2, T_f et on obtient :

	m_1	m_2	T_1	T_2	T_f
X_i	200 g	200 g	20 °C	70 °C	43 °C
$u(X_i)$	2 g	2 g	1 °C	2 °C	2 °C

On peut appliquer la méthode de Monte Carlo pour déterminer $\bar{\mu}$ et $u(\mu)$.

Le code python est le suivant :

```
N = 100000
# valeurs moyennes et incertitude type pour chaque variable
m1 = 200
um1 = 2
m2 = 200
um2 = 2
t1 = 20
ut1 = 1
t2 = 70
ut2 = 2
tf = 43
utf = 2

# calcul du demi intervalle pour utilisation dans la loi uniforme
Deltam1 = um1*3**0.5
Deltam2 = um2*3**0.5
Deltat1 = ut1*3**0.5
Deltat2 = ut2*3**0.5
Deltatf = utf*3**0.5

# tirage aléatoire de N valeurs pour x1 et x2, selon la loi uniforme
M1 = np.random.uniform(low=m1-Deltam1,high=m1+Deltam1,size=N)
M2 = np.random.uniform(low=m2-Deltam2,high=m2+Deltam2,size=N)
T1 = np.random.uniform(low=t1-Deltat1,high=t1+Deltat1,size=N)
T2 = np.random.uniform(low=t2-Deltat2,high=t2+Deltat2,size=N)
Tf = np.random.uniform(low=tf-Deltatf,high=tf+Deltatf,size=N)

#Calcul de l'ensemble des valeurs de d
Mu = (M1*(T1-Tf)+M2*(T2-Tf))/(Tf-T1)

#Calcul de la moyenne et de l'écart type des valeurs issus des N tirages
moy = np.mean(Mu)
```

```
sigma = np.std(Mu,ddof=1)

print('resultats monteCarlo')
print('Moyenne =', moy)
print('Ecart type =', sigma)
```

et conduit aux résultats suivants :

```
resultats monteCarlo
Moyenne = 38.732269491231214
Ecart type = 44.07416412909744
```

On obtient donc :

$$\bar{\mu} = 39 \text{ g}$$

$$u(\mu) = 44 \text{ g}$$

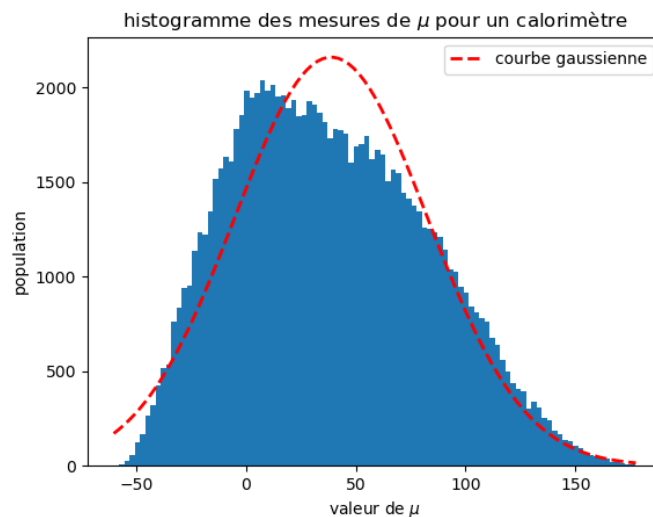
Remarque 1 :

L'incertitude type de cette mesure est très élevée. Le protocole de mesure est peu adapté, il est très probable que lors de l'expérience, l'ensemble des groupes d'un TP trouvent des valeurs très dispersées, certains pouvant même trouver une valeur de μ négative (sans que l'on puisse dire pour autant que ces groupes aient mal travaillé). Le protocole de mesure n'est pas très bon.

Remarque 2 :

Les résultats peuvent être complétés par le tracé de l'histogramme des valeurs simulées, et comparées à la loi normale de paramètres $(\bar{\mu}, u(\mu))$. Le code python suivant réalise le tracé.

```
# Histogramme courbe gaussienne modèle
plt.figure()
plt.title('histogramme des mesures de $\mu$ pour un calorimètre')
n_bins = 100
count, bins, ignored = plt.hist(Mu, bins=n_bins)
plt.plot(bins, N*(bins[-1]-bins[0])/n_bins*1/(sigma * np.sqrt(2 * np.pi)) * np.exp( -
(bins - moy)**2 / (2 * sigma**2) ), '--', linewidth=2, color='r', label='courbe
gaussienne')
plt.legend()
plt.xlabel('valeur de $\mu$')
plt.ylabel('population')
plt.show()
```



On voit notamment apparaître le fait que la loi de probabilité associée à μ n'est pas exactement gaussienne.

Remarque 3 :

Le calcul de $u(\mu)$ a été effectué ici par Méthode de Monte-Carlo. Il est évidemment possible d'appliquer la formule mathématique à partir de l'expression

$$\mu = \frac{m_1(T_1 - T_f) + m_2(T_2 - T_f)}{T_f - T_1} = f(m_1, m_2, T_1, T_2, T_f)$$

et de calculer :

$$u(\mu) = \sqrt{\sum \left(\frac{\partial f}{\partial X_i} \right)_{\bar{\mu}}^2 u(X_i)^2}$$

L'application est assez lourde, et il est dans ce cas préférable d'effectuer une simulation (c'est souvent le cas quand la fonction f possède de nombreuses variables).

5 Comparaison de deux valeurs : z-score

5.1 Acceptabilité d'une mesure individuelle

On considère une variable aléatoire X suivant une loi normale de moyenne μ et d'écart type $u(X) = \sigma$. On peut construire une grandeur appelée

écart centré normalisé, ou z-score :

$$Z = \frac{X - \mu}{u(X)}$$

Par construction, la grandeur z suit donc une loi normale de paramètres $\bar{Z} = 0$, et $u(Z) = 1$. La probabilité qu'une mesure x soit comprise dans l'intervalle $[\mu - 2\sigma, \mu + 2\sigma]$ est égale à la probabilité que z soit compris dans $[-2, 2]$. D'après les résultats relatifs à la loi normale, cette probabilité est d'environ 0,95.

Ceci permet d'établir un critère permettant de décider, avec un certain intervalle de confiance, si une mesure particulière $X = x$ est acceptable (ie si elle est compatible avec les propriétés statistiques de la série) :

$$|z| = \left| \frac{x - \mu}{u(X)} \right| > 2 \Leftrightarrow \text{la valeur } x \text{ n'est pas acceptable}$$

Si X ne suit pas une loi normale, ce résultat n'est pas applicable en toute rigueur. Cependant, on pourra continuer de l'appliquer de manière approchée (ceci est justifié par le fait qu'une grandeur dépendant d'un nombre N de variables aléatoires indépendantes suit une loi de probabilité qui tend vers la loi normale pour N tendant vers l'infini).

Remarque

En assimilant la moyenne $\mu = \bar{x}$ à une valeur x_{ref} , par exemple une valeur tabulée, on peut également écrire $Z = (X - x_{ref})/u(X)$. Cette dernière expression est assez fréquemment utilisée, mais repose sur l'hypothèse $\bar{X} = x_{ref}$. On a alors pour une mesure $X = x$ particulière

$$z = \frac{x - x_{ref}}{u(X)}$$

on obtient alors le critère suivant :

$$|z| = \left| \frac{x - x_{ref}}{u(X)} \right| > 2 \Leftrightarrow \text{la valeur } x \text{ n'est pas acceptable}$$

Représentation graphique associée

Pour une série de mesures, on peut représenter pour chaque valeur de x la valeur de z associée (avec les barres d'erreurs seuils à $2 u(z) = 2$ correspondantes), ce qui permet d'identifier rapidement une mesure suspecte. On peut aussi tracer les valeurs de x (avec les barres d'erreurs seuils à $2 u(x)$ correspondantes)

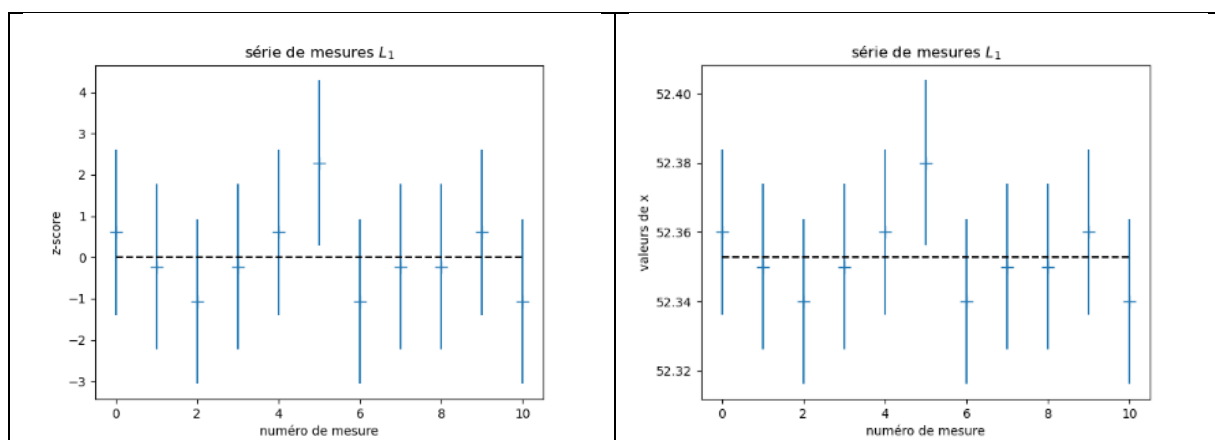
Exemple :

On considère un ensemble de mesures de longueur $X = L_1$ (en cm) :

L_1	52,36	52,35	52,34	52,35	52,36	52,38	52,34	52,35	52,36	52,34
-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------

On suppose connue l'incertitude type $u(X) = 1,0 \times 10^{-2}$, identique pour chaque mesure.

On peut calculer la moyenne $\mu = 52,353$ de la distribution puis représenter les valeurs du z-score pour chaque mesure (avec des barres d'erreur à ± 2) ou bien les valeurs de x (avec des barres d'erreur à $\pm \sigma$). On obtient les deux représentations suivantes :



On voit que toutes les mesures sauf une sont acceptables. Dans ce cas, il faut revenir sur cette mesure : il y a pu avoir une erreur de l'expérimentateur, ou bien $u(X)$ a été sous-estimé.

Le code des tracés précédents est :

```
L1 = np.array([52.36, 52.35, 52.34, 52.35, 52.36, 52.38, 52.34, 52.35, 52.36, 52.34])
mu = np.mean(L1)
sigma = 0.01
```

```

Z = (L1-mu)/sigma
X = [i for i in range(len(L1))]

plt.figure(1)
plt.errorbar(X,Z,yerr=2,fmt='+',markersize = 10)
plt.plot(X,0*Z,'--k')
plt.ylabel('z-score')
plt.xlabel('numéro de mesure')
plt.title('série de mesures $L_1$')

plt.figure(2)
plt.errorbar(X,L1,yerr=2*sigma,fmt='+',markersize = 10)
plt.plot(X,mu*np.ones(L1.shape),'k')
plt.ylabel('valeurs de x')
plt.xlabel('numéro de mesure')
plt.title('série de mesures $L_1$')

plt.show()

```

5.2 Comparaison entre deux mesures

Si l'on cherche à comparer deux mesures X_1 et X_2 d'une même grandeur (par exemple obtenues avec deux protocoles différents, ou par deux groupes d'expérimentateurs différents, on peut de même construire la grandeur :

$$Z = \frac{X_2 - X_1}{\sqrt{u(X_1)^2 + u(X_2)^2}}$$

où $u(X_1)$ et $u(X_2)$ sont les incertitudes types associées à chacune des mesures.

En supposant que X_1 et X_2 aient les mêmes moyennes, on a

$$\bar{Z} = \frac{\bar{X}_2 - \bar{X}_1}{\sqrt{u(X_1)^2 + u(X_2)^2}} = 0$$

$$u(Z)^2 = \left(\frac{1}{\sqrt{u(X_1)^2 + u(X_2)^2}} \right)^2 (u(X_1)^2 + u(X_2)^2) = 1$$

En considérant en plus que X_1 et X_2 suivent des lois normales, on en déduit que Z suit une loi normale de moyenne nulle et d'écart type 1.

En appliquant le même critère, on peut en déduire que pour deux mesures $X_1 = x_1$ et $X_2 = x_2$,

$$|z| = \left| \frac{x_2 - x_1}{\sqrt{u(x_1)^2 + u(x_2)^2}} \right| > 2 \Leftrightarrow \text{les valeurs } x_1 \text{ et } x_2 \text{ ne sont pas compatibles}$$

Remarque

Si l'on cherche à établir la validité d'une mesure $X = X_2$ vis-à-vis d'une grandeur tabulée x_{ref} dont on connaît l'incertitude type $u(x_{ref})$, alors le z-score s'écrit

$$Z = \frac{X_2 - x_{ref}}{\sqrt{u(x_{ref})^2 + u(X_2)^2}}$$

Pour $u(x_{ref})^2 \ll u(X_2)^2$, cette expression peut se simplifier en

$$Z \approx \frac{X_2 - x_{ref}}{\sqrt{u(X_2)^2}} = \frac{X - x_{ref}}{u(X)}$$

On retrouve ainsi l'expression du paragraphe précédent.

6 Accord expérience Modèle - Régression linéaire

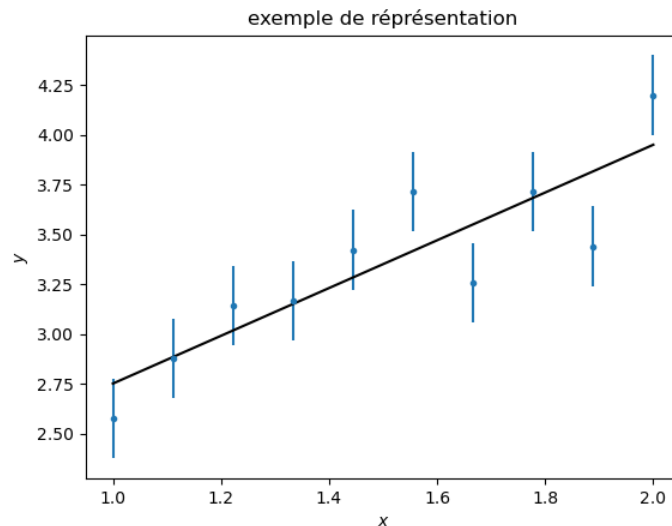
On est souvent confronté en sciences à la problématique suivante : un ensemble de mesures expérimentales est-il compatible avec un modèle théorique ? Autrement dit, pour une grandeur y supposée dépendre d'une grandeur x via une loi théorique $y = f(x)$, les mesures effectuées sont-elles en accord avec la loi théorique ?

6.1 Représentation graphique

On suppose ici que l'on dispose d'un ensemble de N couples de points de mesure (x_i, y_i) , $i \in \{1, \dots, N\}$, avec pour chaque point, des incertitudes $u(x_i)$ et $u(y_i)$.

Dans le cas où on a $u(x_i) \ll u(y_i)$, chaque mesure peut être représentée par une barre verticale centrée sur (x_i, y_i) , et de hauteur $\pm 2u(y_i)$. On peut alors visualiser qualitativement si une certaine corrélation existe entre les valeurs x et y .

On obtient la représentation typique suivante



6.2 Régression linéaire et méthode des moindres carrés

On se restreint dans cette étude à un modèle de type affine $f(x) = ax + b$. L'objectif est de déterminer les « meilleures » valeurs de a et b associées aux données expérimentales. On parle alors de régression linéaire.

6.2.1 Un peu de théorie

On construit une fonction d'erreur, destinée à mesurer l'écart des données expérimentales à la loi modèle :

$$Err = \sum_{i=1}^N (y_i - f(x_i))^2 = \sum_{i=1}^N (y_i - (ax_i + b))^2$$

Cette fonction $Err(a, b)$ est minimale pour les valeurs de a et b vérifiant les conditions¹ :

$$\frac{\partial Err}{\partial a} = 0 \text{ et } \frac{\partial Err}{\partial b} = 0$$

On obtient ainsi

$$\begin{aligned} \sum_{i=1}^N x_i (y_i - (ax_i + b)) &= 0 \\ \sum_{i=1}^N (y_i - (ax_i + b)) &= 0 \end{aligned}$$

soit

$$\begin{aligned} \sum_{i=1}^N x_i y_i - a \sum_{i=1}^N x_i^2 - b \sum_{i=1}^N x_i &= 0 \\ \sum_{i=1}^N y_i - a \sum_{i=1}^N x_i - Nb &= 0 \end{aligned}$$

On peut en déduire les expressions de a et b qui minimisent l'erreur :

$$a = \frac{\overline{xy} - \bar{x}\bar{y}}{\overline{x^2} - \bar{x}^2}, \quad b = \bar{y} - a\bar{x}$$

6.2.2 Instructions python associées

La régression linéaire peut être réalisée dans le langage python à l'aide d'un ajustement polynomial de degré 1, à l'aide de l'instruction `np.polyfit`.

Si $X = [x_1, \dots, x_N]$ et $Y = [y_1, \dots, y_N]$ sont deux tableaux numpy ou deux listes, `np.polyfit(X, Y, 1)` renvoie un tableau numpy unidimensionnel contenant les valeurs $[a, b]$ recherchées (dans cet ordre).

6.3 Validité de la modélisation

Pour chaque mesure (x_i, y_i) , on a un écart $res_i = y_i - f(x_i)$ faible mais a priori non nul ce qui fait que les points de mesure ne sont pas sur la droite modèle d'équation $y = ax + b$.

On peut alors superposer chaque point de mesure accompagné de sa barre d'erreur au tracé de la droite moyenne obtenue. Les barres d'incertitude étant parfois trop petites pour pouvoir être visualisées sur le graphique, on peut en complément tracer le résidu res_i pour mesure.

Exemple 1 : indice n en fonction de λ

Prenons l'exemple de la dépendance de l'indice n d'un milieu donc la valeur dépend de la longueur d'onde λ de la lumière utilisée.

¹ En réalité cette condition est juste une condition d'extrémalité. Il faudrait pousser l'analyse pour vérifier que l'extremum obtenu est bien un minimum.

$\lambda(nm)$	404.7	435.8	480.0	546.1	578.1	615
n	1.77610	1.76325	1.74957	1.73481	1.72942	1.72397
$u(n)$	0.00014	0.00014	0.00014	0.00013	0.00013	0.00013

Le modèle théorique prévoit pour l'indice $n(\lambda)$ une dépendance du type (loi de Cauchy)

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

du type

$$y = ax + b$$

avec $y = n$ et $x = 1/\lambda^2$.

On peut donc écrire les lignes suivantes :

```
lamb = np.array([404.7, 435.8, 480.0, 546.1, 578.1, 615])
n = np.array([1.77610, 1.76325, 1.74957, 1.73481, 1.72942, 1.72397])
p = np.polyfit(1/lamb**2, n, 1)
```

qui donne pour p les valeurs suivantes

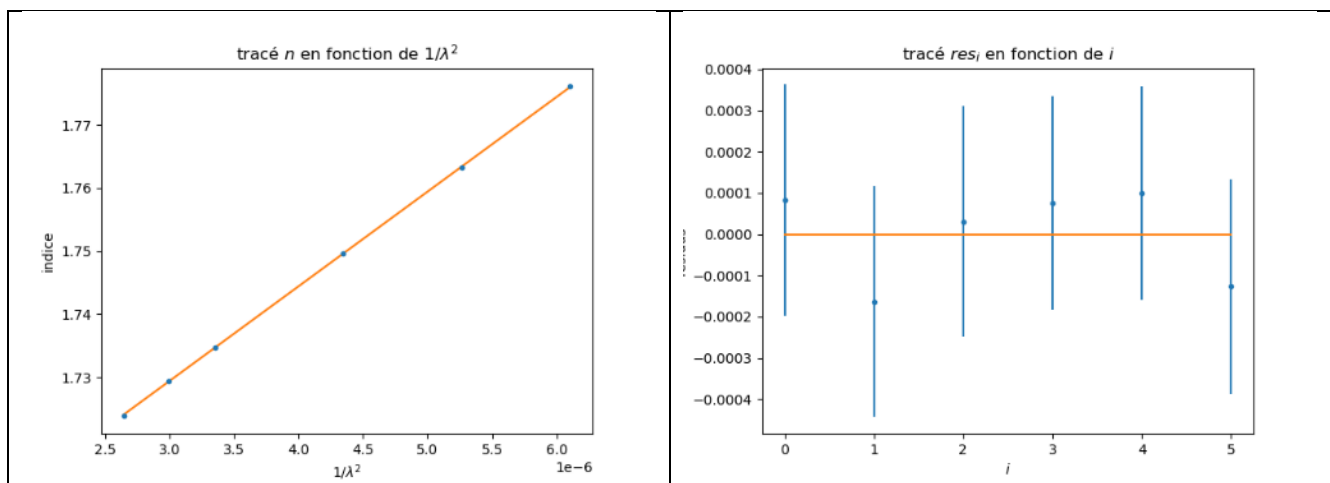
```
>>> p
array([1.49984419e+04, 1.68444157e+00])
```

On obtient donc ici les valeurs :

$$a = B = 1.49984419 \times 10^4 \text{ nm}^2$$

$$b = A = 1.68444157$$

Le tracé de $n = f(1/\lambda^2)$ et des résidus fournit mes deux graphes :



On observe ici que les résidus sont distribués aléatoirement de part et d'autre de la valeur 0, et que les barres d'incertitude à $\pm 2 u(y_i)$ contiennent cette valeur 0. On peut en déduire que les mesures sont compatibles avec le modèle linéaire associé aux valeurs de a et b obtenues par la méthode des moindres carrés.

Le code associé aux deux tracés précédents est

```
inv_lamb = np.linspace(1/lamb[-1]**2, 1/lamb[0]**2, 200) # abscisses de tracé
droite_reg = np.polyval(p, inv_lamb)

plt.figure()
plt.errorbar(1/lamb**2, n, yerr=2*un, fmt='.')
plt.plot(inv_lamb, droite_reg)
plt.ylabel('indice')
plt.xlabel('$1/\lambda^2$')
plt.title('tracé $n$ en fonction de $1/\lambda^2$')

plt.figure()
plt.errorbar(num, n - (p[0]/lamb**2 + p[1]), yerr=2*un, fmt='.')
plt.plot(num, 0*num)
plt.ylabel('résidus')
plt.xlabel('$i$')
plt.title('tracé $res_i$ en fonction de $i$')

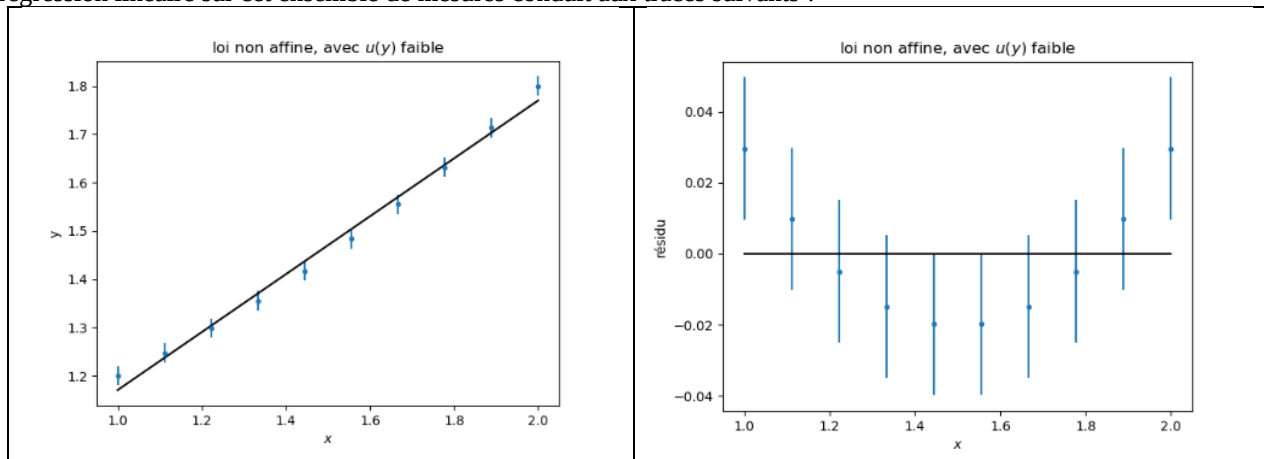
plt.show()
```

Exemple 2 : Autre série de mesures

Prenons l'ensemble de couples de points fictifs (x, y) suivants :

x	1	1.1	1.2	1.3	1.4	1.5	1.6	1.7	1.8	1.9	2.0
y	1.2	1.242	1.288	1.338	1.392	1.45	1.512	1.578	1.648	1.722	1.8
$u(y)$	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01

Une régression linéaire sur cet ensemble de mesures conduit aux tracés suivants :



On observe ici que les résidus ne semblent pas distribués aléatoirement, et que les barres d'incertitudes ne contiennent pas la valeur 0. Les points expérimentaux ne sont pas en accord avec un modèle de type linéaire.

Coefficient de corrélation

Les logiciels de traitement de données fournissent souvent un coefficient de corrélation linéaire, défini par

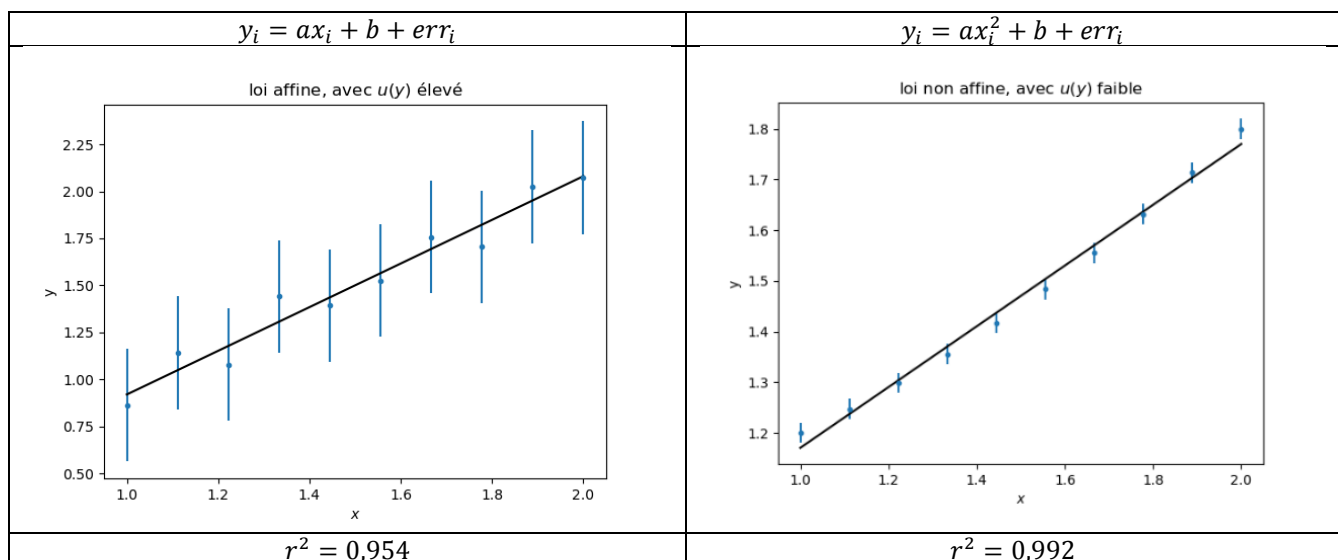
$$r = \frac{\overline{xy} - \bar{x}\bar{y}}{u(x)u(y)}$$

La valeur de r^2 est souvent utilisée, à tort, pour valider l'accord des mesures avec un modèle linéaire (on $r^2 \in [0,1]$, la proximité de r^2 avec 1 étant citée comme critère de validation). Ceci est *faux en général*, comme le montrent les exemples suivants, tracés pour des points de mesure fictifs.

Pour une première série de points, on a créé, à partir de deux variables aléatoires X et Err (qui représente un bruit aléatoire), une variable Y définie par $Y = aX + b + Err$, où a et b sont des réels, et telle que $u(X) = 0$ et $u(Err) = 0,15$. Les variables X et Y sont par construction corrélées de manière affine, mais fortement bruitées.

Pour une deuxième série de points, On a créé selon le même principe une variable Y définie par $Y = aX^2 + b + Err$, avec $u(X) = 0$ et $u(Err) = 0,01$. Ici les variables X et Y ne suivent pas une loi affine.

On obtient pour chaque loi un ensemble de points (x_i, y_i) , pour lesquels on peut appliquer la méthode de régression linéaire et calculer le coefficient r^2 . On obtient le résultat suivant, où il apparaît que la proximité de r^2 avec 1 n'est pas un critère suffisant pour valider l'hypothèse de dépendance linéaire de Y avec X .



6.4 Incertitude sur les paramètres a et b .

Une régression linéaire obtenue à partir d'un jeu de données (x_i, y_i) conduit à un couple de valeurs a et b réalisant la modélisation de la variable y par la droite d'équation $y = ax + b$.

Cependant, les valeurs (x_i, y_i) étant connues avec des incertitudes type $u(x_i)$ et $u(y_i)$, les valeurs obtenues pour a et b sont elles-mêmes des variables aléatoires dépendant de l'expérience utilisée (une autre expérience aurait conduit à un autre jeu de valeurs (x_i, y_i)). On peut donc chercher à déterminer les grandeurs $\bar{a}, \bar{b}$ et $u(a), u(b)$ associées à a, b .

Pour cela on va effectuer une approche aléatoire de type Monte Carlo, et on va se placer dans le cas où $u(x_i) \ll u(y_i)$.

La méthode est la suivante :

- Fixer un nombre N_{exp} de simulations d'expérience
- Pour chaque simulation k
 - tirer aléatoirement des valeurs de $\{y_i\}$ en utilisant une loi de probabilité
 - Déterminer les valeurs a_k et b_k de la régression linéaire sur les valeurs simulées
- Déterminer moyenne et écart-type des $\{a_k\}$ et $\{b_k\}$, et représenter éventuellement leur distribution (histogramme)

En reprenant les données sur la loi de Cauchy (dépendance $n(\lambda)$) et en utilisant l'algorithme précédent, on obtient :

```
# 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
optiques
u_n = np.array([1.4, 1.4, 1.4, 1.3, 1.3, 1.3])*1e-4 # incertitudes-type sur les
indices

# Simulation Monte-Carlo

NMC = 50000 # nombre d'expériences simulées
AMC, BMC = [], [] # listes des valeurs de A et B correspondant à chaque expérience
simulée

for i in range(NMC): # pour chaque expérience
# on simule la mesure des n pour les 6 longueurs d'onde
nMC = n + np.random.normal(0, u_n, size = 6)
# on reprend l'ajustement affine avec cette série de valeurs de n
p = np.polyfit(1/lamb**2, nMC, 1)
# on stocke les valeurs des paramètres d'ajustement dans les listes AMC et BMC
AMC.append(p[1])
BMC.append(p[0])
# résultats de simulation
Amoy = np.mean(AMC)
u_A = np.std(AMC, ddof = 1)

Bmoy = np.mean(BMC)
u_B = np.std(BMC, ddof = 1)

plt.figure(figsize = (15, 4))
plt.subplot(1, 2, 1)
plt.hist(AMC, bins = 'rice', color = 'y')
plt.xlabel(r"Valeurs de $A$ obtenues")
plt.ylabel(r"Nombre d'apparitions")
plt.subplot(1, 2, 2)
plt.hist(BMC, bins = 'rice', color = 'm')
plt.xlabel(r"Valeurs de $B$ obtenues")
plt.ylabel(r"Nombre d'apparitions")
plt.show()
```

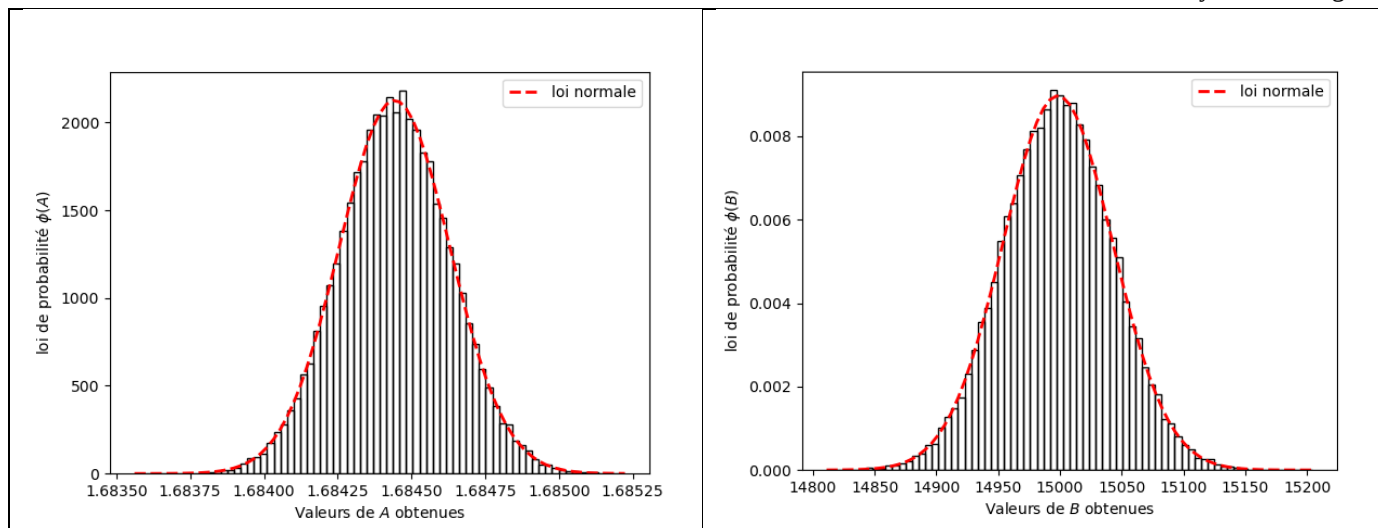
L'exécution conduit à

$$\bar{A} = 1,68444, \quad u(A) = 1,9 \times 10^{-4}$$

$$\bar{B} = 14998 \text{ nm}^2, \quad u(B) = 44 \text{ nm}^2$$

On retrouve ainsi les valeurs de A et B obtenues lors de la régression linéaire effectuée sur un seul jeu de données, cas traité en 6.3 (exemple1), mais on obtient en plus une estimation des incertitudes type sur les coefficients A et B .

On peut compléter cette étude avec le tracé des distributions statistiques pour les valeurs de A et B , où on voit que ces variables suivent des lois de type gaussien.



Annexes

7 Estimateurs de valeur moyenne et écart type

Pour une variable aléatoire X , on peut introduire les variables aléatoires :

$$X^* = \frac{1}{N} \sum_{i=0}^{N-1} X_i$$

et

$$S_{exp}^2 = \frac{1}{N-1} \sum_{i=0}^{N-1} (X_i - X^*)^2$$

pour lesquelles les différents X_i correspondent à des tirages indépendants de la variable X .
On a par linéarité de la valeur moyenne

$$\overline{X^*} = \frac{1}{N} \sum_{i=0}^{N-1} \overline{X_i} = \frac{1}{N} \sum_{i=0}^{N-1} \overline{X} = \overline{X}$$

Par ailleurs, on peut montrer que

$$\overline{S_{exp}^2} = u(X)^2$$

En ce sens, on peut dire que grandeurs X^* et S_{exp}^2 sont des estimateurs non biaisés des grandeurs $\overline{X}$ et $u(X)$.