

Mesure et incertitude

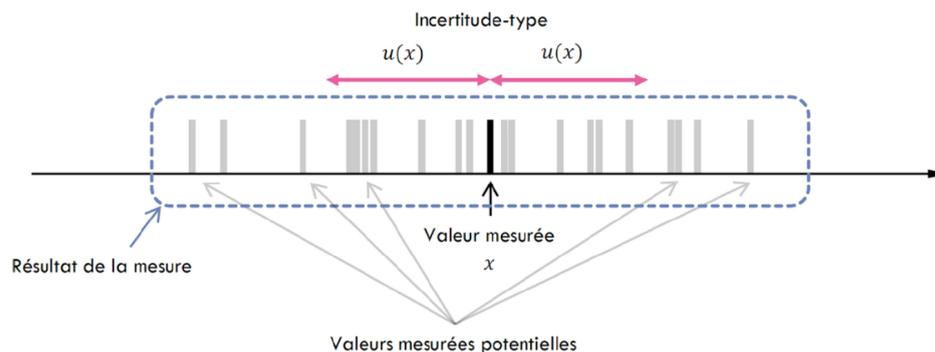
Lord Kelvin écrivait « *il n'y a de science que du mesurable...* ». Mesurer des grandeurs identifiées est une activité fondamentale dans les laboratoires de recherche scientifique et dans l'industrie. Toute validation théorique d'un phénomène (physique, biologique, chimique, etc.) passe par la mesure fiable de ses effets. C'est aussi fondamental dans de nombreuses activités quotidiennes. Les capteurs et instruments ont pour vocation d'aider à traduire la valeur d'une grandeur en une donnée numérique, parfois très simplifiée, qui pourra être utilisée afin de tirer des conclusions ou prendre des décisions, c'est-à-dire en une donnée numérique digne de confiance. Les méthodes et conventions qui régissent la définition, l'évaluation et l'expression des résultats de mesure et les propriétés des instruments sont partie intégrante du langage à vocation universelle de la métrologie, science de la mesure. C'est en particulier grâce à une estimation convenable de l'incertitude attachée à un résultat que ce dernier pourra, paradoxalement, inspirer une confiance maîtrisée, et être reconnu sans équivoque par plusieurs partenaires. Comment, en effet, comparer entre eux des résultats de façon fiable, ou confronter une donnée à une tolérance, en l'absence de caractérisation de l'incertitude ?

Mesurer une grandeur, c'est donc d'une part comparer une grandeur inconnue à une grandeur prise comme référence (unité) à l'aide d'une chaîne instrumentale, et d'autre part estimer une incertitude attachée à cette mesure, afin d'en qualifier la qualité.

1 Variabilité d'une mesure d'une grandeur physique

La mesure parfaite n'existe pas, il s'agit d'un processus complexe qui conduit naturellement à une variabilité de celle-ci. Cette variabilité peut être liée à l'opérateur (deux personnes réalisant la même expérience dans les mêmes conditions aboutiront à des valeurs différentes), à l'environnement (influence des conditions de température sur une expérience longue par exemple), aux instruments de mesure, à la méthode de mesure ...

Cette variabilité est riche d'enseignements sur le processus de mesure. Dès lors, on considèrera que le résultat d'une mesure n'est pas une valeur unique mais un ensemble de valeurs numériques qu'on peut raisonnablement attribuer à la grandeur mesurée.



L'incertitude-type $u(x)$ indique la dispersion des valeurs au sein de l'ensemble {Résultat de la mesure} et rend compte de la variabilité de la mesure. Elle peut être évaluée à l'aide d'une méthode statistique (via la notion d'écart-type) ou non statistique.

2 Evaluation statistique d'une incertitude : incertitude-type de type A

L'évaluation statistique de l'incertitude repose par définition sur la répétition d'une mesure.

Imaginons mesurer la résistance d'un conducteur ohmique, alimenté par une source de tension continue variable, à l'aide d'un ampèremètre et d'un voltmètre.

On mesure $U = 1,02$ V et $I = 2,13$ mA. On en déduit, grâce à la loi d'Ohm : $R = \frac{U}{I} = \frac{1,02}{2,13 \times 10^{-3}} = 479 \Omega$

Pour estimer la variabilité de cette mesure, il faut pouvoir calculer un écart-type. On doit donc reproduire l'expérience.

Une série de $n = 10$ observations pourrait donner :

Observation	1	2	3	4	5	6	7	8	9	10
U (V)	1,02	2,01	3,05	4,03	4,93	5,97	7,01	8,07	9,02	9,96
I (mA)	2,13	4,27	6,36	8,48	10,67	12,81	14,88	17,03	19,27	21,33

On calcule l'ensemble des valeurs expérimentales de la résistance qu'on représente sous forme d'un histogramme (qui rend compte de la variabilité de la mesure) à l'aide d'un programme Python :

```
import numpy as np          # pour manipuler des tableaux
import matplotlib.pyplot as plt # pour les représentations graphiques

### Données expérimentales

U = np.array([1.02,2.01,3.05,4.03,4.93,5.97,7.01,8.07,9.02,9.96]) # remplit un tableau avec
                                                                    les valeurs de U observées
                                                                    (en V)

I = np.array([2.13,4.27,6.36,8.48,10.67,12.81,14.88,17.03,19.27,21.33]) # remplit un tableau
                                                                    avec les valeurs de I
                                                                    observées (en mA)

### Calcul des valeurs de la résistance

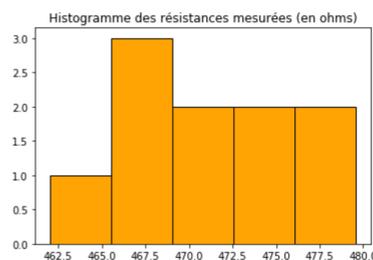
R = U / (I*1e-3)          # calcule la valeur de R (en ohms) pour chaque couple (U,I) et stocke
                                                                    les résultats dans un tableau

### Tracé de l'historgramme - visualisation de la variabilité de la mesure

plt.hist(R, bins='rice', color = 'orange', edgecolor='black') # trace l'historgramme des
                                                                    valeurs observées. La commande
                                                                    'rice' optimise les intervalles
                                                                    d'affichage.

plt.title("Histogramme des résistances mesurées (en ohms)")
plt.show
```

Le programme renvoie :



L'écart-type s_X de cette série de valeurs représente leur dispersion et exprime la variabilité d'une observation. Il constitue ainsi l'incertitude-type associée à une observation unique que l'on note et est défini par :

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

où X_i est la valeur pour la i -ème observation et \bar{X} la moyenne des valeurs calculées telle que

$$\bar{X} = \frac{1}{n} \sum_{i=1}^n X_i$$

On peut compléter le programme :

```
import numpy as np          # pour manipuler des tableaux
import matplotlib.pyplot as plt # pour les représentations graphiques

#%% Données expérimentales

U = np.array([1.02,2.01,3.05,4.03,4.93,5.97,7.01,8.07,9.02,9.96]) # remplit un tableau avec
                                                                # les valeurs de U observées
                                                                # (en V)

I = np.array([2.13,4.27,6.36,8.48,10.67,12.81,14.88,17.03,19.27,21.33]) # remplit un tableau
                                                                # avec les valeurs de I
                                                                # observées (en mA)

#%% Calcul des valeurs de la résistance

R = U / (I*1e-3)          # calcule la valeur de R (en ohms) pour chaque couple (U,I) et stocke
                          # les résultats dans un tableau

#%% Tracé de l'histogramme - visualisation de la variabilité de la mesure

plt.hist(R, bins='rice', color = 'orange', edgecolor='black') # trace l'histogramme des
                                                                # valeurs observées. La commande 'rice' optimise les intervalles d'affichage.
plt.title("Histogramme des résistances mesurées (en ohms)")
plt.show

#%% Calcul de l'incertitude-type

u_R = np.std(R, ddof = 1) # calcule l'écart-type expérimental (la commande ddof = 1 sert à
                          # tenir compte de la division par n-1 dans la formule de
                          # l'écart-type)

print("L'incertitude-type sur une mesure de la résistance vaut {:.2f} ohms".format(u_R))
```

Par convention, on donne l'incertitude-type, elle-même incertaine, à 2 chiffres significatifs.

Dans notre exemple, on retiendra donc $u(R) = 5,7 \Omega$.

Ainsi, pour l'observation initiale, on écrira $R = 479,0 \pm 5,7 \Omega$. La précision de l'incertitude impose le nombre de chiffres significatifs à retenir pour le résultat (précision au dixième d'ohm ici).

Mais souvent, disposant de plusieurs observations, on s'intéresse à la moyenne, \bar{R} , comme meilleur estimateur de la grandeur mesurée. En effet, la moyenne compense les observations les plus élevées par les plus basses. Elle est plus juste qu'une observation unique.

Il reste à estimer la variabilité (et donc l'incertitude) de cette moyenne unique. Une première solution consisterait à reproduire fois l'expérience précédente (série de n observations). On calculerait à chaque fois la moyenne obtenue puis on étudierait la dispersion de l'ensemble obtenu : l'écart-type de cet ensemble de moyennes est par définition l'incertitude-type associée à une moyenne unique. Mais ce processus est coûteux en temps et en ressources. On peut alors avantageusement utiliser le théorème mathématique : $u(\bar{R}) = \frac{u(R)}{\sqrt{n}}$ qui fournit l'écart-type sur la moyenne de n observations.

Dans notre exemple, $\bar{R} = 471,25 \Omega$ (l'incertitude-type n'étant pas encore calculée, on ne peut pas arrondir le résultat) et $u(\bar{R}) = \frac{5,7}{\sqrt{10}}$ pour $n = 10$ observations. Alors, on aura $R = 471,3 \pm 1,8 \Omega$.

3 Evaluation non statistique d'une incertitude : incertitude-type de type B

Lorsqu'une évaluation statistique (de type A) n'est pas possible ou souhaitée, on se tourne vers d'autres méthodes qui comportent une part d'arbitraire puisqu'il va falloir reconstruire la variabilité de la mesure.

Reprenons l'exemple précédent mais cette fois-ci, on ne dispose pas d'un voltmètre et d'un ampèremètre mais seulement d'un ohmmètre. L'appareil de mesure n'étant pas très résolu, il y a de grandes chances d'observer plusieurs fois la même valeur si on répète la mesure. Et répéter n fois la même mesure présente donc ici peu d'intérêt.

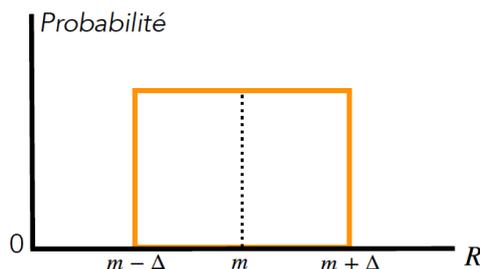
Au cours d'une observation unique, on mesure $R = 470,6 \Omega$. On doit alors estimer l'incertitude type de façon théorique en s'appuyant sur les données fournies par le constructeur. Par exemple, pour l'ohmmètre utilisé, la

notice indique : sur le calibre 500 Ω, précision 0,5 % de la valeur lue + 3 digits. 1 digit correspond à la plus petite valeur lue sur le calibre choisi : 0,1 Ω dans le cas présent.

On peut alors définir un intervalle de valeurs dans lequel on est raisonnablement certain de trouver le résultat de mesure m .

Ici, $m = 470,6 \text{ } \Omega$ et $\Delta = 0,005 \times 470,6 + 3 \times 0,1 = 2,7 \text{ } \Omega$ (précision de l'appareil).

On fait ensuite l'hypothèse que le résultat de mesure est décrit par une variable aléatoire à densité uniforme entre deux bornes (distribution rectangulaire) :



On est certain que $R \in [m - \Delta, m + \Delta]$. Toutes les valeurs de cet intervalle sont équiprobables.

Dans le cas d'une loi de probabilité uniforme, on en déduit l'incertitude-type sur R grâce à la formule : $u(R) = \frac{\Delta R}{\sqrt{3}}$.

Ici, on aurait $u(R) = \frac{2,7}{\sqrt{3}} = 1,5 \text{ } \Omega$ et $R = 470,6 \pm 1,5 \text{ } \Omega$

4 Composition des incertitudes

Reprenons le cas de la toute première observation : mesure d'une résistance à l'aide d'un ampèremètre et d'un voltmètre. On mesure $U = 1,02 \text{ V}$ et $I = 2,13 \text{ mA}$. On en déduit, grâce à la loi d'Ohm : $R = \frac{U}{I} = \frac{1,02}{2,13 \times 10^{-3}} = 479 \text{ } \Omega$.

Si par manque de temps par exemple, on ne peut pas procéder à une évaluation statistique de l'incertitude-type, il faut :

- déterminer l'incertitude-type sur U et I par une évaluation de type B.
- déterminer l'incertitude-type sur R par composition des incertitudes (permet de comprendre comment les incertitudes sur U et I influencent celle sur R).

Les notices des multimètres indiquent :

- Pour le voltmètre : précision 0,3 % de la valeur lue + 2 digits (1digit = 0,01 V sur le calibre 2 V). On en déduit $\Delta_u = 1,02 \times 0,003 + 2 \times 0,01 = 0,023 \text{ V}$ et donc $u(U) = \frac{\Delta_u}{\sqrt{3}} = 0,013 \text{ V}$.

- Pour l'ampèremètre : précision 0,3 % de la valeur lue + 2 digits (1digit = 0,001 mA sur le calibre 2 mA). On en déduit $\Delta_I = 2,13 \times 0,003 + 2 \times 0,001 = 0,0084 \text{ mA}$ et donc $u(I) = \frac{\Delta_I}{\sqrt{3}} = 0,0048 \text{ mA}$.

L'incertitude-type sur se déduit de l'une des relations suivantes :

Cas	Relation	Incertitude
1	$G = aX$ (a constante)	$u(G) = a \times u(X)$
2	$G = X \pm Y$	$u(G) = \sqrt{(u(X))^2 + (u(Y))^2}$
3	$G = \frac{X}{Y}$ ou $G = X \times Y$	$u(G) = G \sqrt{\left(\frac{u(X)}{X}\right)^2 + \left(\frac{u(Y)}{Y}\right)^2}$
4	$G = \lambda X^a Y^b$ (λ, a, b constantes)	$u(G) = G \sqrt{a^2 \left(\frac{u(X)}{X}\right)^2 + b^2 \left(\frac{u(Y)}{Y}\right)^2}$

La loi d'Ohm donne $R = \frac{U}{I}$ donc on se trouve dans le cas 3. On aura alors :

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

$$\text{A.N. : } u(R) = 479 \times \sqrt{\left(\frac{0,013}{1,02}\right)^2 + \left(\frac{0,0048}{2,13}\right)^2} = 6,2 \Omega.$$

Et finalement, $R = 479,0 \pm 6,2 \Omega$.

Pour des cas de figure plus compliqués, la détermination de l'incertitude par le calcul peut s'avérer longue et difficile.

On pourra alors utiliser une simulation Monte Carlo pour remédier au problème. Explicitons la méthode sur l'exemple de la mesure de la résistance :

- Connaissant une mesure m_U de U et Δ_U , on simule un grand nombre de valeurs de tension en considérant comme une variable aléatoire suivant : une distribution uniforme centrée sur m_U et de précision Δ_U .
- De même, connaissant une mesure m_I de I et Δ_I , on simule un grand nombre de valeurs de courant en considérant comme une variable aléatoire suivant une distribution uniforme centrée sur m_I et de précision Δ_I .
- On calcule autant de valeurs de R que de valeurs de U et I simulées : la valeur mesurée de R correspond alors à la moyenne des valeurs simulées.
- On calcule enfin l'écart-type associé à l'ensemble des valeurs qui nous donne l'incertitude $u(R)$.

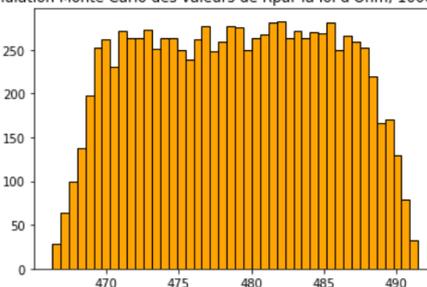
Le programme Python, avec une boucle, pourrait ressembler à :

```
import numpy as np
import matplotlib.pyplot as plt
#%%Entrées et incertitudes sur les entrées
U = 1.02 # Tension mesurée (en V)
I = 2.13 # Courant mesuré (en mA)
Delta_U = 0.023 # Précision sur la tension (en V)
Delta_I = 0.0084 # Précision sur le courant (en mA)
#%%Simulation Monte Carlo
N = 10000 # Nombre de simulations
R = [] # Liste vide pour stocker les valeurs de R simulées
for i in range(N): # Boucle qui va répéter N fois les instructions
    tension = np.random.uniform(U-Delta_U, U+Delta_U) # Tire aléatoirement une valeur de la
    courant = np.random.uniform(I-Delta_I, I+Delta_I) # Tire aléatoirement une valeur du courant
    resistance = tension / (courant * 1e-3) # Calcule la résistance par la loi d'Ohm
    R.append(resistance) # Ajoute la valeur de la résistance à la liste R
Rmoy = np.mean(R) # Calcule la moyenne des N valeurs de la résistance
u_R = np.std(R, ddof = 1) # Calcule l'écart-type expérimental sur une valeur de R
#%% Résultats et histogramme des valeurs simulées de R
print("La valeur mesurée de R est: {:.1f} ohms.".format(U / (I*1e-3)))
print("L'incertitude sur R est: {:.1f} ohms.".format(u_R))
plt.hist(R, bins='rice', color = 'orange', edgecolor = 'black')
plt.title("Simulation Monte Carlo des valeurs de R par la loi d'Ohm, 10000 essais")
plt.show
```

Le programme renvoie :

```
La valeur mesurée de R est: 478.8 ohms.
L'incertitude sur R est: 6.3 ohms.
```

Simulation Monte Carlo des valeurs de R par la loi d'Ohm, 10000 essais



On retrouve un résultat en accord avec celui obtenu par l'application de la formule de composition des incertitudes et on a $R = 478,8 \pm 6,3 \Omega$.

5 Comparaison de deux valeurs : écart normalisé ou z-score

Il s'agit ici de déterminer si deux valeurs (soit deux résultats de mesure x_1 et x_2 , soit un résultat de mesure et une valeur de référence x_{ref}) sont compatibles en réalisant un test statistique.

- Comparaison de x_1 et x_2 : connaissant les incertitudes $u(x_1)$ et $u(x_2)$ sur et on calcule le z-score (ou écart normalisé) :

$$z_{12} = \frac{|x_1 - x_2|}{\sqrt{(u(x_1))^2 + (u(x_2))^2}}$$

x_1 et x_2 sont compatibles si $z_{12} < 2$

- Comparaison de x et x_{ref} : connaissant l'incertitude $u(x)$ sur x et considérant que la valeur de référence est connue avec certitude ($u(x_{ref}) = 0$), on calcule le z-score ou écart-normalisé :

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

x et x_{ref} sont compatibles si $z < 2$.

Le seuil $z = 2$ correspond à une probabilité de 5 % de se tromper sous l'hypothèse de deux valeurs compatibles. Autrement dit, au-dessus du seuil $z = 2$, l'évènement « les deux valeurs sont compatibles » est rare et on décide de rejeter l'hypothèse.

La valeur du seuil est arbitraire et dépend du champ d'application du calcul du z-score. Il a été historiquement introduit par Ronald Fisher, biologiste et statisticien britannique qui a révolutionné l'application des statistiques aux sciences au début du 20ème siècle.

Champ d'application	Seuil
TP	2
Médecine, psychologie, sociologie, économie	2
Physique des particules, années 1960	3
Physique des particules, aujourd'hui	5

6 Regression linéaire

Comment peut-on tester l'adéquation d'un modèle avec des résultats expérimentaux ?

Dans le cas d'une regression linéaire, on suppose qu'il existe une fonction affine $y = ax + b$ reliant deux variables x et y du modèle, c'est-à-dire en la recherche d'une droite passant au plus près des N points expérimentaux (x_i, y_i) . On cherche alors à déterminer le coefficient directeur a (pente) de la droite et son ordonnée à l'origine b .

On utilise pour ce faire par exemple la méthode « des moindres carrés » en cherchant à minimiser la quantité

$$\sum_{i=1}^n [y_i - (ax_i + b)]^2$$

On obtient alors

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

et

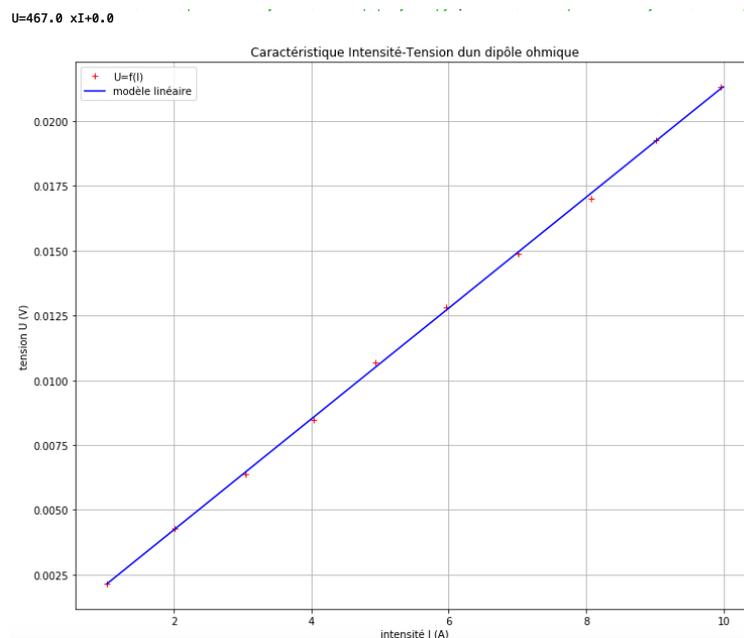
$$b = \bar{y} - a\bar{x}$$

Réalisation avec Python : on peut utiliser la fonction « polyfit » de la bibliothèque « numpy ». Elle nécessite trois arguments : la liste des abscisses, la liste des ordonnées et le degré du polynôme choisi pour le modèle (un ici). Les listes sont déclarées comme des tableaux numpy (« array »).

Reprenons l'exemple de la résistance, on peut utiliser le programme suivant :

```
import numpy as np
import matplotlib.pyplot as plt
U=np.array([1.02,2.01,3.05,4.03,4.93,5.97,7.01, 8.07, 9.02, 9.96])
I=np.array([2.13*1e-3,4.27*1e-3,6.36*1e-3,8.48*1e-3,10.67*1e-3,12.81*1e-3
,14.88*1e-3, 17.03*1e-3, 19.27*1e-3, 21.33*1e-3])
coeff=np.polyfit(I, U,1)
Umodel = coeff[0]*I+coeff[1]
print('U={0:.1f}'.format(coeff[0]),'xI+{0:.1f}'.format(coeff[1]))
fig = plt.figure(figsize=(12,10))
plt.plot(U,I,'r+',label='U=f(I)')
plt.plot(Umodel,I,'b-',label='modèle linéaire')
plt.legend()
plt.xlabel("intensité I (A)")
plt.ylabel("tension U (V)")
plt.grid()
plt.title("Caractéristique Intensité-Tension dun dipôle ohmique")
plt.show()
```

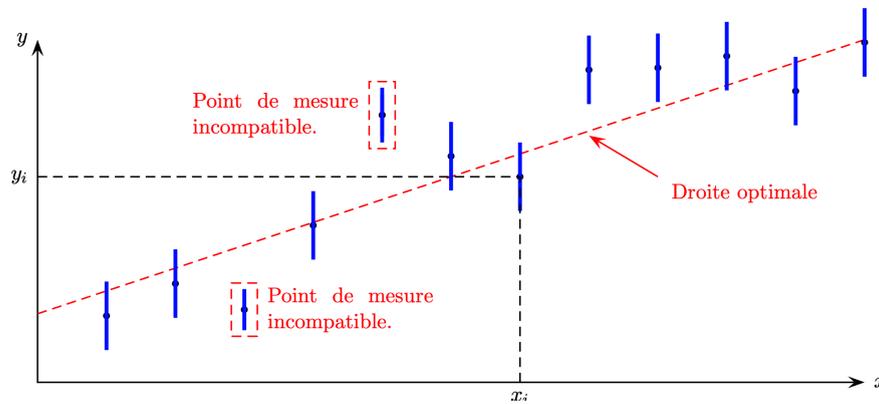
Le programme renvoie :



Pour s'assurer qu'une régression linéaire est correcte, on tracera systématiquement sur un graphique les données mesurées ainsi que la droite de la régression linéaire. Le modèle sera validé si, à l'œil, les points de mesure sont bien alignés et que la droite passe le plus proche de tous les points possibles, en incluant leurs incertitudes-type. On rappelle que l'incertitude-type est une estimation de la variabilité de la mesure. Ainsi, il est naturel que les points expérimentaux soient éloignés de la valeur de la modélisation de quelques incertitudes-types.

On évitera en général l'utilisation du coefficient de corrélation r^2 dont l'interprétation est notoirement délicate. On peut par contre tracer les résidus (écart entre l'ordonnée du point expérimental et la droite de régression) pour confronter visuellement les données aux hypothèses de l'algorithme ; on le fera notamment si on ne voit pas de manière évidente que la droite tracée « passe » par les barres d'incertitudes, parce que celles-ci sont trop petites. On pourra enfin rapporter les résidus à l'incertitude-type qui leur est associée, pour tracer les écarts normalisés.

Dans l'exemple général ci-dessous, le modèle est validé :



Comment déterminer l'incertitude sur les paramètres a et b du modèle ?

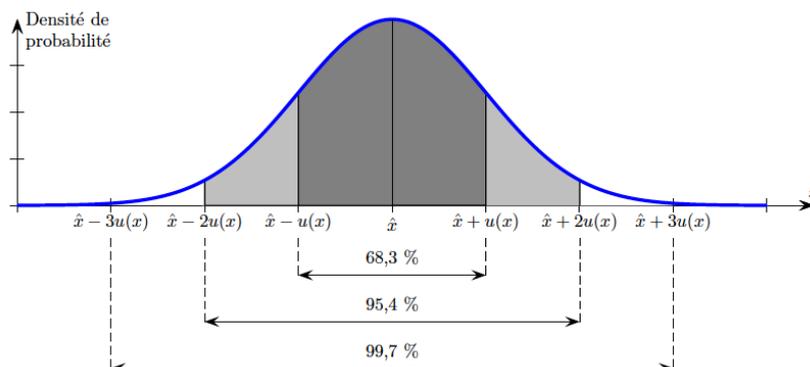
Principe :

On réalise une simulation Monte-Carlo en réalisant un très grand nombre de régressions linéaires sur une série de points sans incertitudes. Pour obtenir ces points, on génère aléatoirement pour chaque point mesuré une valeur à l'aide des incertitudes-types expérimentales. Conformément aux préconisations citées précédemment, si on n'a pas d'information pour savoir de quelle façon générer les valeurs, on choisit une distribution de probabilité uniforme.

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

Annexes

Loi normale ou gaussienne

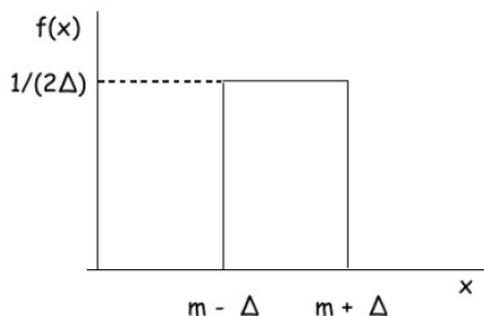


On appelle loi normale $\mathcal{N}(\mu, \sigma^2)$, la loi de probabilité de la variable aléatoire X de densité

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

$\mu = \hat{x}$ est l'espérance mathématique ou moyenne, σ^2 la variance, σ l'écart-type de la variable aléatoire X .

Loi de probabilité uniforme ou rectangulaire



- La probabilité de trouver x dans l'intervalle $[m - \Delta, m + \Delta]$ est égale à 1.
- Δ est l'incertitude maximale
- On peut calculer l'espérance

$$\int_{-\infty}^{+\infty} x \cdot f(x) dx = \int_{m-\Delta}^{m+\Delta} x \cdot \frac{1}{2\Delta} dx = m$$

et la variance :

$$\int_{-\infty}^{+\infty} (x - m)^2 \cdot f(x) dx = \int_{m-\Delta}^{m+\Delta} (x - m)^2 \cdot \frac{1}{2\Delta} dx = \frac{\Delta^2}{3}$$

d'où l'écart-type :

$$\sigma = \frac{\Delta}{\sqrt{3}}$$