



### Ressources pour le cycle terminal général et technologique

---

## Mesure et incertitudes

Ces documents peuvent être utilisés et modifiés librement dans le cadre des activités d'enseignement scolaire, hors exploitation commerciale.

Toute reproduction totale ou partielle à d'autres fins est soumise à une autorisation préalable du Directeur général de l'enseignement scolaire.

La violation de ces dispositions est passible des sanctions édictées à l'article L.335-2 du Code de la propriété intellectuelle.

Mai 2012

## Sommaire

---

<b>Introduction</b>	<b>2</b>
<b>I. Une vision probabiliste de l'erreur</b>	<b>3</b>
A. La notion d'erreur.....	3
1. Un exemple pour commencer.....	3
2. Pourquoi une telle variabilité des résultats ?.....	4
3. La notion d'erreur.....	4
4. Comment traiter de la variabilité : la « randomisation ».....	5
5. Les composantes de l'erreur.....	5
B. L'incertitude.....	8
1. Notion d'incertitude-type.....	8
2. Différents modes d'évaluation de l'incertitude sur une grandeur.....	10
3. Évaluation de type A d'une incertitude-type.....	10
4. Évaluation de type B d'une incertitude-type.....	12
5. Détermination d'incertitudes de type B.....	13
6. Recommandations pratiques.....	13
<b>II. Incertitude-type composée</b>	<b>15</b>
A. Incertitude-type composée sur un mesurage.....	15
B. Détermination de l'incertitude-type composée.....	15
<b>III. Incertitude élargie</b>	<b>17</b>
A. Notion d'incertitude élargie.....	17
1. Incertitude élargie.....	17
2. Détermination du facteur d'élargissement $k$ .....	17
B. Présentation des résultats.....	18
1. Arrondissement.....	18
2. Présentation des résultats.....	19
C. En conclusion.....	19
D. Un exemple.....	20
<b>IV. Annexes</b>	<b>22</b>
Annexe 1 : Les incertitudes-types sur le mesurage d'une grandeur.....	22
Annexe 2 : La Démarche de recherche des causes.....	24
Annexe 3 : Démarche de détermination d'une incertitude sur une grandeur Y.....	25
Annexe 4 : Un rappel des lois de probabilité.....	26
Annexe 5 : Les recommandations de détermination d'incertitude de type B.....	28
Annexe 6 : La loi normale.....	30
Annexe 7 : Incertitude composée.....	35
Annexe 8 : Incertitude sur l'incertitude.....	36
<b>V. Bibliographie</b>	<b>37</b>

## Introduction

---

« Une erreur peut devenir exacte, selon que celui qui l'a commise s'est trompé ou non. »  
Pierre Dac ; Les pensées - Ed. du Cherche Midi (1972)

Ce document a pour vocation de présenter la vision probabiliste de l'erreur, développée depuis environ trois décennies par le Bureau international des poids et mesures (BIPM) et qui a permis d'installer un consensus international dans l'expression de l'incertitude de mesure.

Il se veut être une ressource pour les enseignants de sciences physiques et de mathématiques des lycées.

Pour les enseignants de sciences physiques elle veut donner à comprendre les raisons et les mécanismes mis en œuvre derrière les formules qui sont appliquées dans les estimations de mesures de grandeur, par exemple dans le programme de première STL, [www.education.gouv.fr/cid55406/mene1104103a.html](http://www.education.gouv.fr/cid55406/mene1104103a.html) en complétant ainsi les documents déjà parus : *Nombres, mesures et incertitudes* (Inspection générale Sciences Physiques et Chimiques) <http://eduscol.education.fr/pid23213-cid46456/ressources-pour-le-college-et-le-lycee.html>

Pour les enseignants de mathématiques, elle donnera des exemples d'utilisation des notions probabilistes enseignées au lycée, en particulier en liant la notion d'erreur à celle de variable aléatoire, celle d'incertitude-type avec celle d'écart-type.

Outre la nécessité d'une connaissance partagée sur un sujet qui relève des deux disciplines, ce domaine du calcul d'incertitude devrait donner la possibilité de travaux communs développés par les enseignants de mathématiques et de sciences physiques.

L'ambition reste cependant modeste, notamment dans les outils présentés ; une bibliographie proposée en fin de document donne des références pour ceux qui souhaiteraient approfondir le sujet, en particulier dans l'étude de l'incertitude des mesures obtenues à partir de données corrélées ou appariées ou encore dans le cas de données obtenues en faible nombre.

Vision probabiliste : pourquoi, alors qu'on travaille sur des données statistiques ? Essentiellement parce qu'on est dans une activité de modélisation et que l'on cherche à passer de quelques observations à une caractéristique de l'ensemble de toutes les observations possibles, que des données obtenues on va chercher à induire des connaissances sur des variables aléatoires, qu'à partir d'un nombre fini de données on va estimer une connaissance sur une infinité de possibilités, et en particulier qu'on va donner des renseignements sur des modèles considérés comme continus à partir d'un nombre fini d'observations. Il faut garder présent à l'esprit cette idée de modèle tout au long de ce document.

Cette brochure est conçue pour pouvoir être lue sans être arrêté par des difficultés dans des développements mathématiques qui sont renvoyés en annexe. On y trouvera également quelques compléments qui peuvent éclairer les choix faits.

# I. Une vision probabiliste de l'erreur

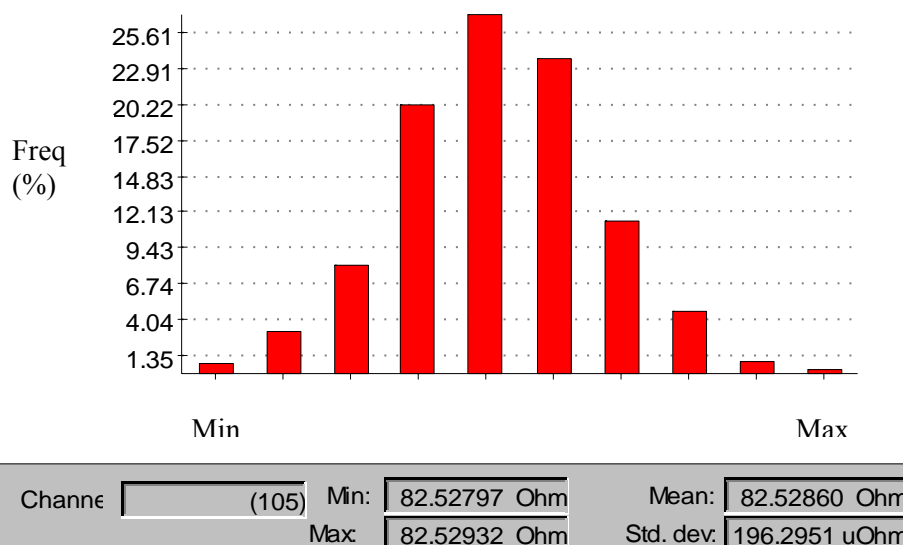
## A. La notion d'erreur

### 1. Un exemple pour commencer

On souhaite mesurer une résistance. Le conducteur ohmique dont on souhaite mesurer la résistance est branché aux bornes d'un ohmmètre. On utilise une première technique de mesure utilisant « quatre fils » de liaison entre le conducteur ohmique et l'instrument.

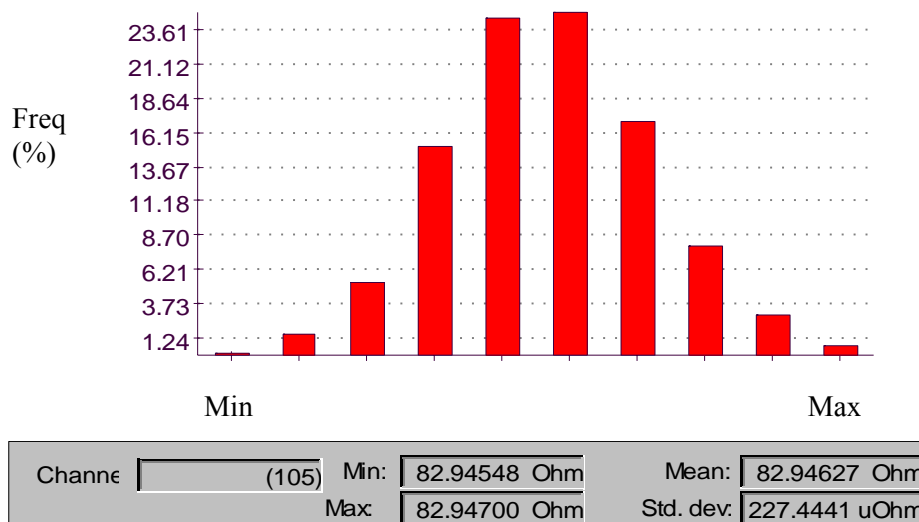
Notre instrument communique avec un ordinateur et l'on utilise un programme d'acquisition de données. Ce programme effectue 2000 mesures  $m$  de la résistance  $R$ , repère les valeurs  $m_{\min}$  et  $m_{\max}$ , divise l'intervalle  $[m_{\min}; m_{\max}]$  en 10 intervalles (classes), calcule le nombre  $n$  de résultats dans chaque classe et affiche les résultats sous la forme d'un diagramme.

On obtient les résultats ci-dessous :



Recommençons la mesure précédente en configurant notre instrument en ohmmètre « deux fils », ce qui correspond à une mesure courante de la valeur d'une résistance.

On obtient désormais les résultats suivants :



Des questions se posent au vu de ces résultats :

1. Pourquoi une telle variabilité des résultats ?
2. Pourquoi ces deux méthodes donnent-elles des résultats différents ?
3. Et enfin, finalement quelle est la mesure de la résistance cherchée ?

On s'était aperçu depuis longtemps, notamment en astronomie, science qui possédait les instruments les plus précis, que :

- plusieurs mesurages d'une même caractéristique donnaient souvent des valeurs différentes,
- la répartition des résultats avait une « forme en cloche » : « *il n'y a aucun doute que les petites erreurs ont lieu plus souvent que les grandes* ».

Toute la problématique de la détermination de la mesure d'une grandeur est là : parmi tous ces résultats lequel choisir et comment estimer l'erreur qui pourrait être commise ?

## **2. Pourquoi une telle variabilité des résultats ?**

Une série de mesures est soumise à des conditions environnementales qui modifient les résultats obtenus :

- la grandeur à mesurer n'est pas parfaitement définie, la largeur d'une table n'est pas un objet défini, la taille d'une pièce métallique dépend de sa position, la surface d'un liquide n'est pas plane, ...
- les conditions environnementales évoluent (température, pression, ...)
- l'instrument de mesure est source d'erreur (temps de réponse, exactitude, sensibilité)
- l'opérateur ne refait jamais la même mesure exactement dans les mêmes conditions (fatigue, erreurs de parallaxe, effet de ménisque dans une pipette...)

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

Précisons quelques termes de vocabulaire du domaine de la métrologie et qui vont être employés :

*Mesurage* : ensemble d'opérations ayant pour but de déterminer une valeur d'une grandeur.

*Mesurande* : grandeur particulière soumise à mesurage (longueur, masse, intensité, ...).

« *Valeur vraie* » d'un *mesurande* : mesure que l'on obtiendrait par un mesurage parfait. On ne la connaît pas et on parle également de « valeur théorique ».

*Grandeur d'influence* : grandeur qui n'est pas le *mesurande* mais qui a un effet sur le résultat du mesurage.

## **3. La notion d'erreur**

Si  $y_i$  est le résultat d'un mesurage et  $y_0$  la « valeur vraie » du *mesurande*, l'erreur sur le résultat  $y_i$  est le nombre  $e_i = y_i - y_0$

Ce concept d'erreur est idéal et les erreurs ne peuvent malheureusement pas être connues exactement.

Dans la problématique qui nous intéresse on va chercher à estimer une valeur  $y_0$  du *mesurande*, et à quantifier l'erreur commise sur cette estimation.

Ainsi, la démarche visée est de fournir, autour du résultat d'un mesurage, un intervalle dont on puisse s'attendre à ce qu'il comprenne une fraction élevée des valeurs qui pourraient raisonnablement être attribuées au *mesurande*.

#### 4. Comment traiter de la variabilité : la « randomisation »

Dans les années 1960, dans les livres de sciences physiques qui présentaient les « Incertitudes des mesures et calculs approchés », on cherche à définir un majorant des erreurs, et on propose des théorèmes comme « l'incertitude absolue d'une somme ou d'une différence est la somme des incertitudes absolues ». Ainsi par exemple, si l'erreur sur la dimension des côtés  $c$  d'un carré est majorée par 1mm, l'erreur sur les dimensions du périmètre  $p$  de ce même carré est majorée par 4mm. On écrit alors par exemple  $c = 12\text{mm} \pm 1\text{mm}$  et on en déduit que  $p = 48\text{mm} \pm 4\text{mm}$

Quelle réalité représentent ces nombres ? Une erreur de 4mm est-elle possible, plausible, probable ? C'est cette idée qui sous tend la présentation qui va être proposée ci-dessous, donnant un aperçu du calcul d'incertitude de mesures tel qu'il se pratique aujourd'hui en laboratoire ou en milieu industriel.

Cependant, contrairement à l'exemple proposé en introduction, il sera rarement possible d'effectuer 2 000 mesures d'une même grandeur et il va falloir trouver un moyen de décrire la distribution des valeurs possibles des résultats d'un mesurage.

Un mesurage comporte en général plusieurs opérations dont chacune peut être source de variabilité. L'objet de l'étude des erreurs est de pouvoir préciser cette variabilité, et une façon de le faire est d'introduire le « hasard », un hasard qui peut résulter de notre ignorance (Dutarte, Piednoir). Une manière d'interpréter les résultats est de passer par une « randomisation », c'est-à-dire qu'**on explique la variabilité de résultats déterministes (il n'y a pas d'aléatoire dans les mesures) comme s'ils étaient des réalisations d'une variable aléatoire.**

Autrement dit, on remplace la notion d'erreur accidentelle par celle d'incertitude aléatoire : la variabilité de la mesure n'est pas un « accident » évitable, mais est inhérente au processus de mesurage si celui-ci est suffisamment sensible. (Robert, Treiner)

Si l'on répète le mesurage, on obtient une série de valeurs  $y_1, y_2, \dots, y_n$  que l'on considère comme les valeurs prises par une variable aléatoire  $Y$  et une série de valeurs  $e_1, e_2, \dots, e_n$  qui sont les erreurs définies sur chacune des observations ; ces valeurs sont considérées comme celles prises par une variable aléatoire  $E$  :

L'erreur de mesure est une variable aléatoire

On peut ainsi modéliser le mesurage par :  $Y = y_0 + E$

**L'hypothèse fondamentale du traitement probabiliste de l'erreur est que la variable  $E$  obéit à une loi de probabilité « bien définie ».**

**L'objet du calcul d'incertitude sera de déterminer :**

- les paramètres de la loi de probabilité de  $E$ .
- un intervalle dont on puisse s'attendre à ce qu'il comprenne une fraction élevée des valeurs qui pourraient raisonnablement être attribuées au mesurande.

#### 5. Les composantes de l'erreur

On envisage traditionnellement qu'une erreur possède deux composantes, à savoir une composante **aléatoire** et une composante **systématique**.

##### a) Composante aléatoire de l'erreur

L'**erreur aléatoire**  $\Delta$  provient des variations temporelles et spatiales **non prévisibles** de grandeurs d'influence. Les effets de telles variations appelés effets aléatoires entraînent des variations pour les observations répétées du mesurande (bien que le mesurage soit effectué dans des conditions aussi constantes que possible).

*L'erreur aléatoire est liée aux conditions opératoires*

Bien qu'il ne soit pas possible de compenser l'erreur aléatoire d'un résultat de mesure, elle peut être réduite en augmentant le nombre d'observations.

### b) Composante systématique de l'erreur

L'**erreur systématique  $\epsilon$**  se produit sur un résultat de mesure à partir d'un **effet reconnu** d'une grandeur d'influence ; cet effet, appelé **effet systématique**, peut être **quantifié** et s'il est significatif par rapport à la précision requise du mesurage, **une correction** est appliquée au résultat.

Si nous reprenons l'exemple de départ, lors de la deuxième série de mesures, les fils de liaison de l'instrument au conducteur ont une résistance  $R_f$ .

Dans ces conditions, l'instrument ne mesure pas  $R$  mais  $R + R_f$ . Chaque mesure  $m_i$  est systématiquement plus grande que la valeur de  $R$  (des fils de 2 m de long et de faible section ont été utilisés).

La valeur moyenne  $\langle m \rangle$  des  $N$  mesures est plus grande de 0,4172... ohms que dans le cas précédent, passant de 82,5286...ohms pour « quatre fils » à 82,9462...ohms pour une méthode « deux fils », cette différence étant due à cette erreur systématique. Les  $N$  mesures  $m_i$  restent dispersées autour de  $\langle m \rangle$  caractérisant l'erreur aléatoire.

Si on utilise cette méthode à deux fils, une correction de 0,4172... ohms sera effectuée sur chaque résultat. Rien ne nous prouve cependant qu'il n'existe pas d'autres erreurs systématiques !

En général, la correction est une opération difficile car elle nécessite une connaissance approfondie du processus de mesure afin d'identifier au mieux les causes d'erreurs puis d'estimer les corrections à apporter.

Il existe de nombreuses sources d'erreurs systématiques, comme par exemple :

- l'effet des grandeurs d'influence (température, pression,...) ;
- l'erreur de justesse des instruments (décalage du zéro par exemple, chronomètre mal calibré,...) ;
- la position de l'objet mesuré ;
- la perturbation due à la présence des instruments d'observation.

Dans la pratique, différentes méthodes sont utilisées pour détecter et évaluer ces erreurs, comme par exemple :

- mesurer la même grandeur avec un instrument différent ;
- mesurer la même grandeur avec des méthodes différentes ;
- mesurer une grandeur étalon (contrôle de la justesse) ;
- mesurer un même mesurande dans des laboratoires différents.

*L'erreur systématique peut être considérée comme une erreur « constante » qui affecte chacune des observations.*

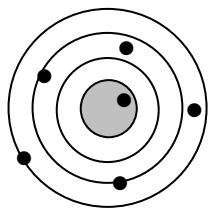
Le plus souvent on aura seulement une majoration de cette erreur « constante ».

L'erreur systématique d'un résultat de mesure ne peut être réduite en augmentant le nombre d'observations, mais par l'application d'une correction.

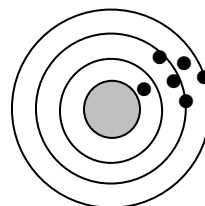
Les corrections étant faites le mieux possible, il subsiste un doute sur la valeur des corrections.

On admet que les variations de l'erreur systématique autour de la correction effectuée sont aléatoires, ce qui permet de supposer que l'erreur systématique  $\epsilon$  suit une loi de probabilité bien définie.

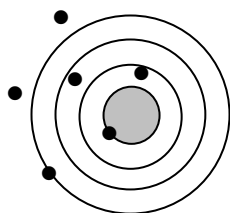
On peut illustrer ces notions d'erreurs systématique et aléatoire par le tir dans une cible :



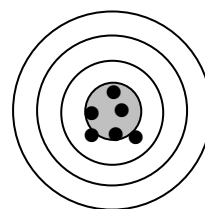
juste, mais pas fidèle  
(valeurs centrées mais dispersées)  
erreurs aléatoires



fidèle, mais pas juste  
(valeurs décentrées mais resserrées)  
erreurs systématiques



ni juste, ni fidèle  
erreurs aléatoires et systématiques



fidèle et juste  
erreurs faibles

Ce dessin n'est cependant qu'une vue théorique trompeuse, car en général, on ne connaît pas la cible, la dispersion nous renseigne sur les erreurs aléatoires, mais la présence d'erreur systématique est souvent difficile à déceler.

### c) Modélisation du mesurage

On suppose que le résultat d'un mesurage a été corrigé pour tous les effets systématiques reconnus comme significatifs et qu'on a fait tous les efforts pour leur identification. On dit alors que la méthode de mesure est **correcte**.

On peut donc modéliser le mesurage par :  $Y = y_0 + \epsilon + \Delta$

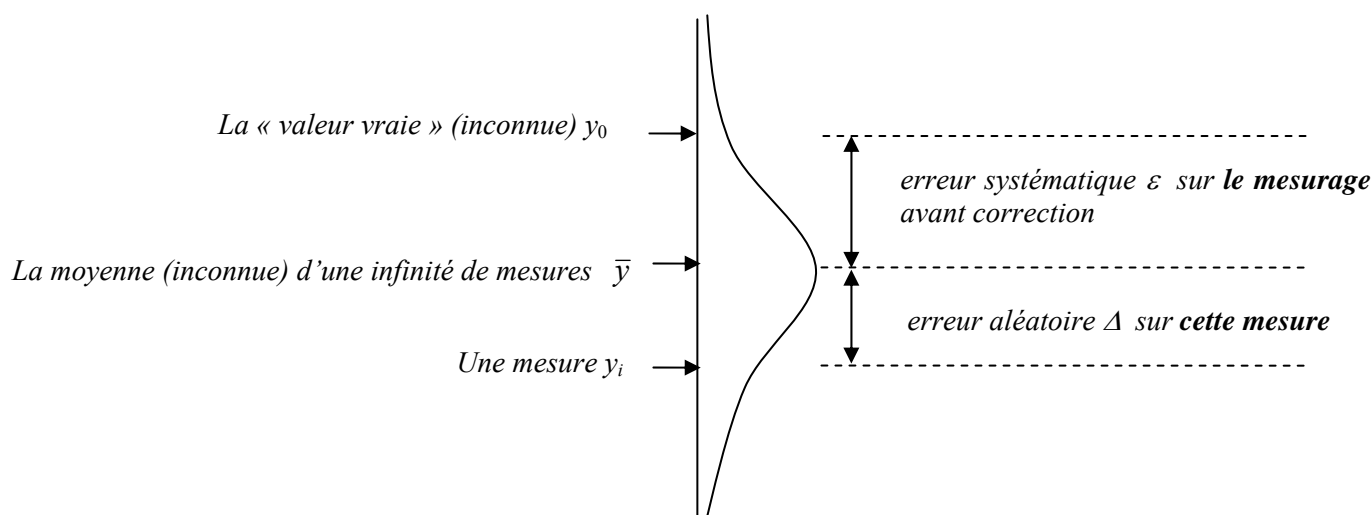
Si on imagine pouvoir faire une infinité de mesures (ce qui revient à considérer la distribution de toutes les mesures), l'**erreur systématique**  $\epsilon$  sur un mesurage est le décalage entre la « valeur vraie » du mesurande et la moyenne (théorique) de l'infinité de toutes les mesures qui pourraient être effectuées.

C'est la « *moyenne qui résulterait d'un nombre infini de mesurages du même mesurande, effectués dans des conditions de répétabilité, moins une valeur vraie du mesurande.* » (VIM 93 ou GUM 08)

Comme on le verra plus loin (ce résultat est justifié en annexe), la moyenne est en général la meilleure estimation de la grandeur mesurée, et l'**erreur aléatoire**  $\Delta$  sur un mesurage représente la différence entre cette moyenne et les résultats obtenus. C'est le « *résultat d'un mesurage moins la moyenne d'un nombre infini de mesurages du même mesurande, effectués dans des conditions de répétabilité.* » (VIM 93)



Le schéma ci-dessous en donne une illustration :



**Une hypothèse pragmatique : il n'y a pas de raison objective pour que les résultats se répartissent plus d'un côté que de l'autre de la « valeur vraie »**

On fera l'hypothèse dans la suite que la méthode de mesure est correcte, ce qui se traduit mathématiquement par :

l'espérance mathématique des variables  $\varepsilon$  et  $\Delta$  est nulle et on a ainsi  $E(Y) = y_0$

Donner une mesure du mesurande va nécessiter la détermination d'une estimation de l'espérance et de l'écart type (ou plus précisément de la variance) de cette variable  $Y$ .

## B. L'incertitude

Le mot incertitude signifie doute ; l'incertitude du résultat d'un mesurage reflète l'impossibilité de connaître exactement la valeur du mesurande.

Dans cette vision probabiliste de l'erreur, le concept d'incertitude est défini en accord avec cette approche :

**Incertitude** : paramètre, associé au résultat d'un mesurage, qui caractérise la dispersion des valeurs qui pourraient être raisonnablement attribuées au mesurande.

### 1. Notion d'incertitude-type

Nous avons vu que le mesurage peut être modélisé par une variable aléatoire  $Y$  d'espérance  $y_0$  et que l'on cherche à caractériser la dispersion des valeurs que peut prendre cette variable aléatoire. Une mesure de cette dispersion peut être obtenue à partir de l'écart-type de la variable aléatoire  $Y$ .

**La détermination de l'incertitude sur le mesurage va être exprimée en fonction de l'écart-type de la variable aléatoire  $Y$ .**

L'écart-type de  $Y$  est appelé **incertitude-type** sur le résultat du mesurage. On note généralement  $u(y)$  cette incertitude-type sur  $Y$ .

**L'essentiel de la démarche va consister à déterminer la loi de probabilité suivie par  $Y$  (ou par  $E$ ) et à estimer la valeur de l'écart-type de  $Y$  (ou de  $E$ ).**

Dans l'exemple de la détermination de la résistance, la donnée des 2 000 résultats permet d'avoir une bonne approximation de la loi de la variable  $R$  (on pourrait vérifier qu'elle est gaussienne) et si on fait l'hypothèse qu'on a maîtrisé les erreurs systématiques, une estimation de l'écart-type donné par le tableau est d'environ  $200\mu\Omega$ .

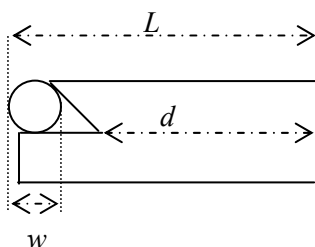
Cependant, il n'est toujours possible d'obtenir une estimation de la grandeur par un recueil de données et  $Y$  peut dépendre de plusieurs autres variables.

Prenons un exemple :

Avant de lancer la production en série, on veut estimer l'incertitude sur la dimension  $d$  en fond de rainure de la pièce ci-dessous.

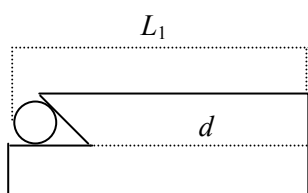


Pour déterminer  $d$ , on place une pige de diamètre  $w$  dans la rainure et on mesure une distance  $L$

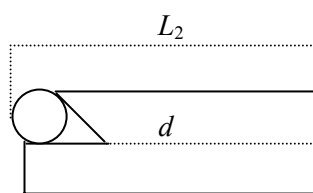


On peut montrer que  $d = L - \frac{w}{2} \left( 1 + \frac{1}{\tan \frac{\alpha}{2}} \right)$  et  $d$  est alors une fonction de  $L$ ,  $w$  et  $\alpha$ .

On peut s'affranchir du contrôle de la mesure de  $\alpha$ , qui n'est pas la plus aisée. Pour cela on effectue deux séries de mesures indépendantes de la longueur  $L$  avec deux pige de diamètres différents ; avec la pige de diamètre  $w_1$ , on obtient une longueur  $L_1$  et avec la pige de diamètre  $w_2$ , une longueur  $L_2$ .



Avec la pige de diamètre  $w_1$



Avec la pige de diamètre  $w_2$

En utilisant la formule déterminée précédemment, on montre alors que  $d = \frac{L_1 w_2 - L_2 w_1}{w_2 - w_1}$

Cette fois  $d$  est une fonction de  $L_1, L_2, w_1$  et  $w_2$ .

On voit apparaître une double difficulté, d'une part l'accès aux valeurs de  $d$  n'est pas direct et d'autre part les données sur les variables dont dépend  $d$  ne sont pas du même ordre ; pour certaines variables,  $L_1, L_2$ , un mesurage va donner un ensemble de données que l'on va pouvoir traiter statistiquement, alors que pour d'autres, comme  $w_1$  et  $w_2$  on accédera à des données proposées par le constructeur.

**Bien que la problématique reste la même, à savoir se donner une idée de la distribution des valeurs que prend  $d$ , évaluer l'incertitude sur  $d$  va demander de combiner deux modes d'évaluation, l'un s'appuyant sur une analyse statistique et l'autre sur une modélisation probabiliste.**

## 2. Différents modes d'évaluation de l'incertitude sur une grandeur

Si on fait varier la totalité des grandeurs dont dépend le résultat d'un mesurage, l'incertitude-type sur chacune des variables peut être évaluée par des moyens statistiques. Cependant, comme cela est rarement possible en pratique, l'incertitude-type sur certaines variables peut être estimée par utilisation d'un modèle probabiliste décrivant la loi de propagation de l'erreur sur cette variable.

Si une grandeur est estimée par des moyens statistiques, on dit qu'on a une évaluation de type A de l'incertitude-type sur cette grandeur.

Si une grandeur est estimée par un modèle probabiliste, on dit qu'on a une évaluation de type B de l'incertitude-type sur cette grandeur.

## 3. Évaluation de type A d'une incertitude-type

On suppose dans ce cas que la grandeur  $X$  est estimée à partir d'une série statistique (on établit par exemple une série de 15 mesures de la longueur d'une pièce)

On notera  $(X_1, X_2, \dots, X_n)$  un  $n$ -échantillon de  $X$ , où  $X_i$  représente la variable aléatoire associée à la  $i^{\text{ème}}$  mesure de la grandeur  $X$ . Les  $n$  mesures  $x_1, x_2, \dots, x_n$  constituent un échantillon des valeurs prises par  $X$ . La variable aléatoire  $\bar{X} = \frac{1}{n} \sum_{i=1}^n X_i$  a pour espérance celle de  $X$  et la moyenne arithmétique des valeurs  $x_1, x_2, \dots, x_n$  est en général une bonne estimation de l'espérance  $E(X)$  de la variable  $X$ .

On prend donc, en général, comme estimation ponctuelle de  $X$  le nombre  $\bar{x}$  défini par :

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

Dans l'exemple du calcul de résistance, on prend 82,5286...ohms (on réserve le débat sur les chiffres significatifs pour plus tard) comme estimation de la résistance mesurée.

Pour une grandeur  $X$  estimée à partir de  $n$  observations répétées indépendantes (ce qui signifie que les variables  $X_i$  sont indépendantes), obtenues dans les mêmes conditions de mesure,  $x_1, x_2, \dots, x_n$  (ce qui impliquerait que chaque opération de mesurage nécessite le démontage et le remontage du dispositif de mesure !), le nombre  $s^2(X) = \frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})^2$  est la « meilleure » estimation de la variance de  $X$  notée  $\sigma^2(X)$ .

$s(X)$  représente une estimation de la dispersion  $\sigma(X)$  des valeurs prises par  $X$  autour de la moyenne  $E(X)$ .

$s(X)$  est appelé écart-type expérimental d'une mesure ou écart-type de répétabilité.

Un résultat classique de statistique sur les lois d'échantillonnage montre que la « meilleure » estimation sur  $\sigma^2(\bar{X})$ , variance de la moyenne  $\bar{X}$  de  $X$ , est  $s^2(\bar{X}) = \frac{s^2(X)}{n}$

L'écart-type expérimental de la moyenne  $s(\bar{X})$  est utilisé comme estimation de l'incertitude de la moyenne  $\bar{X}$ .

En conclusion :

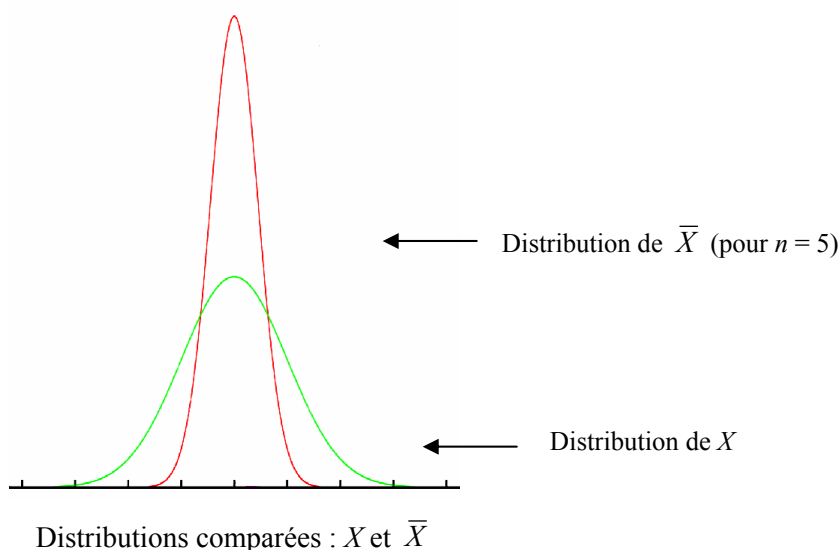
Si une grandeur  $X$  est estimée à partir de  $n$  observations répétées indépendantes  $x_1, x_2, \dots, x_n$ , alors

$$l'incertitude-type  $u(x)$  sur son estimation  $\bar{x} = \frac{x_1 + x_2 + \dots + x_n}{n}$  est  $s(\bar{X}) = \frac{s(X)}{\sqrt{n}} = \frac{1}{\sqrt{n}} \sqrt{\frac{\sum_{i=1}^n (x_i - \bar{x})^2}{(n-1)}}$$$

Il faut bien comprendre ce que signifie ce résultat, c'est une des difficultés souvent rencontrée :

$\sigma(X)$  est le paramètre qui caractérise la dispersion des valeurs prises par  $X$  et caractérise l'incertitude sur **une** mesure. Si on effectue  $n$  mesures,  $\bar{x} = \frac{x_1 + \dots + x_n}{n}$  est une moyenne parmi d'autres possibles (si on refaisait à nouveau  $n$  mesures, on obtiendrait une autre moyenne) ; c'est un résultat dans la population de l'infinité de moyennes possibles.

La théorie des probabilités montre que la distribution de l'ensemble des moyennes est bien moins dispersée (la dispersion est divisée par  $\sqrt{n}$ ) que l'ensemble des mesures uniques. Cela confirme l'idée « pragmatique » que l'estimation à partir d'une moyenne est « meilleure » que sur une mesure seule.



Cependant dans la pratique, pour des problèmes de coût, on ne pourra effectuer  $n$  mesures et on aura souvent une estimation de  $s(X)$ , produite dans des conditions similaires par un autre opérateur ou à un autre moment que celui de l'évaluation de l'incertitude. On fait l'hypothèse que les mesures antérieures constituent une bonne image de la dispersion des mesures attachées à la procédure employée.

La formule précédente n'est donc pas applicable.

On considère alors que l'estimation de  $s(X)$  reste l'écart-type expérimental d'une mesure. Si on effectue  $p$  nouvelles observations indépendantes, alors on prendra  $u(x) = \frac{s(X)}{\sqrt{p}}$

Si une grandeur  $X$  est estimée à partir de  $p$  observations répétées indépendantes  $x_1, x_2, \dots, x_p$  et si  $s(X)$  est l'écart-type expérimental d'une mesure (obtenu auparavant à partir de  $n$  valeurs) alors

$$l'incertitude-type  $u(x)$  sur son estimation  $\bar{x} = \frac{x_1 + x_2 + \dots + x_p}{p}$  est  $\frac{s(X)}{\sqrt{p}}$$$

*Exemple* : On effectue 20 mesures du diamètre d'un cylindre à l'aide d'un pied à coulisse et on obtient  $s(X) = 0,018$  mm

Si l'on effectue l'évaluation de l'incertitude à partir de ces 20 observations, l'incertitude-type retenue sur la moyenne de ces 20 observations sera  $u(x) = \frac{0,018}{\sqrt{20}}$  mm

Si on estime que cette évaluation, 0,018 mm, représente convenablement l'écart-type de la dispersion d'une mesure de ce cylindre autour de sa moyenne, on peut prendre  $s(X) = 0,018$  mm comme l'écart-type expérimental d'une mesure établie à l'aide du même type d'instrument. C'est cette valeur qui sera utilisée pour une nouvelle mesure.

L'incertitude-type sur une mesure sera alors  $u(x) = 0,018$  mm

L'incertitude-type sur la moyenne de trois mesures ultérieures sera alors  $u(x) = \frac{0,018}{\sqrt{3}}$  mm

*Une remarque* : on voit sur cet exemple l'ambiguïté qui peut surgir de cette notation  $u(x)$  qui est employée pour désigner un écart-type représentant une incertitude-type sur la grandeur  $X$ , mais qui dépend de la procédure utilisée. D'où la nécessité, comme nous le verrons dans les exemples traités de bien préciser les conditions du mesurage.

Reprenons l'exemple traité de la dimension en fond de rainure.

On effectue les deux séries de dix mesures de manière indépendante et on obtient, en mm :

<b>L1</b>	52,36	52,35	52,34	52,35	52,36	52,34	52,35	52,35	52,36	52,34
<b>L2</b>	59,17	59,18	59,17	59,17	59,19	59,18	59,18	59,17	59,18	59,19

L'écart-type de répétabilité obtenu sur une mesure de  $L_1$  est égal à  $8,164 \times 10^{-3}$  mm et celui sur une mesure de  $L_2$  est égal à  $7,888 \times 10^{-3}$  mm.

Pour obtenir l'écart-type retenu **sur la moyenne** des 10 mesures de  $L_1$ , on prendrait cette même valeur  $8,164 \times 10^{-3}$  mm divisée par  $\sqrt{10}$ .

#### 4. Évaluation de type B d'une incertitude-type

Lorsque l'estimation d'une grandeur  $X$  ne peut être obtenue à partir d'observations répétées, la variance estimée  $u^2(x)$  ou l'incertitude-type  $u(x)$  sont évaluées par un jugement fondé sur des lois de probabilité supposées a priori.

La détermination de la loi de l'erreur est liée à la maîtrise du processus de mesure et à l'expérience de l'opérateur ; elle dépend d'un ensemble d'informations qui peuvent être :

- des résultats de mesures antérieures ;
- l'expérience ou la connaissance du comportement et des propriétés des matériaux et instruments utilisés ;
- de facteurs d'influence (température, pression,.....) ;
- des spécifications du fabricant ;
- les données fournies par des certificats d'étalonnage ou autres ;
- l'incertitude assignée à des valeurs de référence et donnée avec ces valeurs.

En annexe 4, on trouvera des exemples de lois utilisées dans les calculs d'incertitude ainsi que les incertitudes-types correspondantes. Les lois qui sont rencontrées le plus souvent sont les lois normales et les lois rectangulaires (ou uniformes).

La loi normale ou gaussienne (détaillée en annexe 6) est centrale dans la théorie des erreurs. On constate que le plus souvent, la distribution des erreurs aléatoires est normale (on ne peut pas le montrer mais on peut le vérifier par des tests de normalité), et que même si ce n'est pas le cas, la distribution des moyennes l'est généralement.

Dans l'exemple des mesures de la résistance, la forme du diagramme peut laisser penser raisonnablement que la distribution de l'ensemble des valeurs est normale.

La loi rectangulaire ou uniforme est utilisée souvent en calcul d'incertitude, lorsqu'on ne connaît qu'une majoration de l'erreur, ce qui est souvent le cas pour les erreurs systématiques.

Ainsi si on peut raisonnablement faire l'hypothèse que les erreurs se situent entre deux nombres  $a$  et  $b$ , la loi rectangulaire sur  $[a, b]$  (elle vaut 1 entre  $a$  et  $b$  et 0 ailleurs) est de toutes les lois définies sur ce même intervalle  $[a, b]$ , celle qui a le plus grand écart-type ; pour cela on la nomme parfois la « loi du pire » en ce sens qu'on ne minimise pas l'écart-type qui caractérise l'incertitude-type.

## 5. Détermination d'incertitudes de type B

Le travail va être de choisir, en fonction des informations recueillies ou des connaissances des processus, la loi de probabilité qui lui semble être le mieux représentative du phénomène étudié.

Ainsi, si on sait raisonnablement que les valeurs de la grandeur  $X$  sont comprises entre  $M - d$  et  $M + d$ , le choix de la loi de propagation de  $X$  entre  $M - d$  et  $M + d$  va décider de l'incertitude-type retenue :

Si on suppose que la loi est normale on prendra  $u(x) = \frac{d}{3}$

Si on suppose que la loi est triangulaire, on prendra  $u(x) = \frac{d}{\sqrt{6}}$

Si on suppose que la loi est rectangulaire on prendra  $u(x) = \frac{d}{\sqrt{3}}$

On remarque sur les résultats précédents que l'hypothèse d'une loi triangulaire est un bon compromis entre les lois normales et rectangulaires.

En annexe 5, on trouvera les dispositions qui, sans information supplémentaire, sont prises dans la pratique.

## 6. Recommandations pratiques

### a) **Choix des composantes de l'incertitude**

Dans la pratique il existe de nombreuses sources possibles d'incertitude dans un mesurage, telle que :

- la définition incomplète du mesurande ;
- un échantillonnage non représentatif ;
- une connaissance insuffisante ou un mesurage imparfait des conditions d'environnement ;
- un biais (dû à l'observateur) dans la lecture des instruments analogiques ;
- la résolution de l'instrument ;
- des valeurs inexactes des étalons et matériaux de référence ;
- des valeurs inexactes des constantes et paramètres retenus (obtenus par des sources extérieures par exemple) ;
- des approximations dans la méthode de calcul des incertitudes ou dans le processus de mesure.

Ces sources ne sont pas nécessairement indépendantes, et certaines contribuent aux variations entre les observations répétées du mesurande dans des conditions identiques.

Il est important de ne pas compter deux fois les mêmes composantes de l'incertitude. Si une composante de l'incertitude provenant d'un effet particulier est obtenue par une évaluation de type B, elle ne doit être introduite comme composante indépendante dans le calcul de l'incertitude finale du résultat de mesure que dans la limite où l'effet ne contribue pas à la variabilité des observations répétées.

*Par exemple*, la résolution d'un instrument de mesure, de sensibilité adaptée au mesurage, contribue à la variabilité des observations répétées, par la précision des résultats obtenus. Par contre une erreur éventuelle de justesse de cet instrument n'y contribue pas.

### b) Incertitude-type sur une grandeur

Le mesurage d'une grandeur  $Y$  peut être modélisé par  $Y = y_0 + E_1 + E_2 + \dots + E_n$  où les variables  $E_1, E_2, \dots, E_n$  représentent les différentes composantes **indépendantes** de l'erreur.

Un résultat statistique montre que  $u^2(Y) = u^2(E_1) + u^2(E_2) + \dots + u^2(E_n)$

Par exemple :

Plaçons nous dans le cas où  $Y$  représente le mesurage d'une pièce dans des conditions d'environnement contrôlées. On suppose qu'on effectue une série d'observations à l'aide d'un instrument de mesure et que les composantes retenues de l'erreur amènent au calcul de :

- $u_A$ , incertitude-type déterminée statistiquement sur la série des observations ;
- $u_B$  incertitude-type déterminée sur la justesse de l'instrument de mesure.

Alors l'incertitude retenue sur la grandeur  $Y$  est :  $u^2(Y) = u_A^2 + u_B^2$

Dans l'exemple « fond de rainure » traité, si les mesures de longueurs sont effectuées à l'aide d'un pied à coulisse au 1/100 dont l'erreur de justesse maximale est de  $30\mu\text{m}$ , alors l'incertitude sur une mesure de  $L_1$  est :

- $u_A = 0,00816$  mm et
- $u_B = \frac{0,030}{\sqrt{3}}$  mm (instrument vérifié)

On a alors  $u^2(L_1) = u_A^2 + u_B^2 = 0,00036658 \dots$  mm et on en déduit que  $u(L_1) = 0,0190 \dots$  mm  
(Nous précisons les règles d'arrondissement en fin de document)

*Remarque :*

L'instrument de mesure est adapté au mesurage, les erreurs attachées à la résolution de l'instrument sont prises en compte dans la variabilité des résultats des mesures.

Sinon il faudrait prendre en compte l'erreur attachée à la résolution de l'instrument de mesure. Pour s'en convaincre, il suffit d'imaginer des mesures prises avec un instrument peu précis (comme par exemple un mètre de charpentier) ; toutes les valeurs seraient identiques et l'écart-type de répétabilité égal à zéro !

## II. Incertitude-type composée

Reprenons l'exemple de la détermination des incertitudes sur la dimension en fond de rainure (proposé page 8) ; on cherche à évaluer l'incertitude sur  $d$  défini par  $d = \frac{L_1 w_2 - L_2 w_1}{w_2 - w_1}$ ,  $d$  est une fonction

de  $L_1, L_2, w_1$  et  $w_2$  ; il va falloir déterminer cette incertitude en fonction des incertitudes sur  $L_1, L_2, w_1$  et  $w_2$ , c'est-à-dire l'écart-type de la variable aléatoire modélisant  $d$ .

C'est l'objet de ce paragraphe.

### A. Incertitude-type composée sur un mesurage

En général pour déterminer l'incertitude de mesure associée à un résultat, il faut décrire le processus de mesure et déterminer le modèle mathématique qui relie la valeur mesurée  $y$  du mesurande  $Y$  aux différentes grandeurs qui interviennent dans le processus :

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

où les  $X_i$  sont des grandeurs mesurées, des corrections d'erreurs systématiques, des constantes physiques, des grandeurs d'influence estimées,....

On connaît les lois de répartition des erreurs sur chacune des grandeurs  $X_i$  et donc l'incertitude-type  $u(x_i)$  sur chacune de ces grandeurs. L'objet de ce paragraphe est de déterminer l'écart-type sur la variable  $Y$  qui représentera l'incertitude-type sur  $Y$ , appelée incertitude-type composée de  $Y$  et notée  $u_c(y)$

### B. Détermination de l'incertitude-type composée

**Dans le cas général**, les erreurs étant considérées comme « petites » devant les valeurs des grandeurs, on utilise la formule ci-dessous (dont on trouvera une justification en annexe 7)

Si  $Y = f(X_1, X_2, \dots, X_n)$  avec  $X_1, X_2, \dots, X_n$  indépendantes alors on prend généralement

$$u_c^2(y) = \sum_{i=1}^n \left( \frac{\partial f}{\partial x_i} \right)^2 u^2(x_i)$$

où  $u(x_i)$  est l'incertitude-type sur  $x_i$  et  $u_c(y)$  est l'incertitude-type composée sur  $y$

L'expression ci-dessus s'écrit également  $u_c^2(y) = \sum_{i=1}^n (c_i u(x_i))^2$  avec  $c_i = \frac{\partial f}{\partial x_i}(\mu)$  où  $\mu$  représente la valeur de référence pour chacune des variables (une donnée ou une moyenne de données).

La variance composée  $u_c^2(y)$  peut ainsi être considérée comme une somme de termes dont chacun représente la variance estimée de la contribution de chaque variable  $x_i$ .

**Nous verrons, dans la présentation des résultats, que cette écriture permet de repérer le poids relatif de chacun des facteurs intervenant dans l'estimation de l'incertitude et ainsi prendre des décisions quant aux actions à mettre en œuvre pour diminuer cette incertitude.**

Dans le cas de l'exemple de la détermination de la dimension  $d$  en fond de rainure, on a vu que

$$d = \frac{L_1 w_2 - L_2 w_1}{w_2 - w_1}$$



On obtient :

$$c_1 = \frac{\partial d}{\partial L_1} = \frac{w_2}{w_2 - w_1} = 3, \quad c_2 = \frac{\partial d}{\partial L_2} = \frac{-w_1}{w_2 - w_1} = -2,$$

$$c_3 = \frac{\partial d}{\partial w_1} = \frac{-(L_2 - L_1)w_2}{(w_2 - w_1)^2} = -5,121 \text{ et enfin } c_4 = \frac{\partial d}{\partial w_2} = \frac{(L_2 - L_1)w_1}{(w_2 - w_1)^2} = 3,414$$

En prenant  $w_1 = 8\text{mm}$ ,  $w_2 = 12\text{mm}$ , pour valeur de  $L_1$  la moyenne 52,35mm des 10 valeurs et pour valeur de  $L_2$  la moyenne 59,178mm des 10 valeurs.

Et par conséquent  $u_c^2(d) = (3u(L_1))^2 + (2u(L_2))^2 + (5,121u(w_1))^2 + (3,414u(w_2))^2$

Le calcul explicite de l'incertitude composée sur  $d$  sera effectué en fin du document.

*Remarque 1 :*

Le cas où les variables ne sont manifestement pas indépendantes est nettement plus délicat à traiter. C'est le cas par exemple lorsqu'on détermine à l'aide d'un même multimètre, intensité et différence de potentiel ; il est alors nécessaire d'introduire la covariance entre les couples de variables non indépendantes. On pourra alors se reporter au guide ISO sur le calcul d'incertitude qui proposera des formules à appliquer.

Dans le cas où les incertitudes sont estimées à partir d'observations, un moyen de vérifier si les variables  $X$  et  $Y$  sont indépendantes, sans avoir recours au coefficient de corrélation, est de comparer les écarts-types de  $X + Y$  et  $X - Y$  ; lorsque les résultats sont très différents, on ne peut pas considérer les variables indépendantes.

Sur l'exemple des séries de mesures de  $L_1$  et  $L_2$  :

<b>L1</b>	52,36	52,35	52,34	52,35	52,36	52,34	52,35	52,35	52,36	52,34
<b>L2</b>	59,17	59,18	59,17	59,17	59,19	59,18	59,18	59,17	59,18	59,19
<b>L1 + L2</b>	111,53	111,53	111,51	111,52	111,55	111,52	111,53	111,52	111,54	111,53
<b>L1 - L2</b>	-6,81	-6,83	-6,83	-6,82	-6,83	-6,84	-6,83	-6,82	-6,82	-6,85

On a alors  $\text{var}(L_1 - L_2) = 0,01135292$  et  $\text{var}(L_1 + L_2) = 0,01135293$

On peut remarquer que les deux variances sont très proches, ce qui incite à penser que les variables  $L_1$  et  $L_2$  sont indépendantes (ce qui est assez naturel car la détermination des deux dimensions nécessite un montage et démontage des piges).

*Remarque 2 :* si  $Y$  est de la forme  $Y = kX_1^{p_1} X_2^{p_2} \dots X_n^{p_n}$ , un calcul simple des dérivées partielles montre qu'à partir de la formule déterminant  $u_c(y)$ , on peut écrire une relation déterminant l'incertitude-type composée relative  $\frac{u_c(y)}{|y|}$  en fonction des incertitudes-types composées relatives

$$\frac{u(x_i)}{|x_i|}. \text{ On obtient alors : } \left(\frac{u_c(y)}{y}\right)^2 = \sum_{i=1}^n (p_i \frac{u(x_i)}{x_i})^2 .$$

### III. Incertitude élargie

---

#### A. Notion d'incertitude élargie

##### 1. Incetitude élargie

Rappelons la problématique développée :

Dans les chapitres précédents, on a modélisé la mesure d'une grandeur  $Y$  comme variable aléatoire, et on a déterminé une approximation de l'écart-type de cette variable que l'on a noté  $u_c(y)$ .

L'intention de départ est de fournir, autour du résultat d'un mesurage, un intervalle dont on puisse s'attendre à ce qu'il contienne une fraction élevée de la distribution des valeurs qui pourraient raisonnablement être attribuées au mesurande  $Y$ .

Après l'estimation de l'écart-type de  $Y$ , il reste à estimer la loi de probabilité suivie par cette variable.

**Idéalement, on aimerait déterminer un nombre  $k$  tel que si  $Y$  est estimé par  $y$  avec une incertitude  $U(y) = k u_c(y)$ , alors on peut affirmer que :  $y - U(y) \leq Y \leq y + U(y)$  avec une probabilité  $p$  proche de 1.**

*$U(y)$  notée également  $U$  est appelée incertitude élargie sur  $Y$*

##### 2. Détermination du facteur d'élargissement $k$

La détermination de  $k$  correspond à ce qu'on appelle en statistique la détermination d'un intervalle de confiance à un niveau de confiance  $p$ .

Pour obtenir ce facteur  $k$ , il est nécessaire d'avoir une connaissance de la loi de probabilité de la variable représentée par le résultat de mesure.

Dans la pratique nous n'avons au mieux qu'une estimation de cette loi et de l'écart-type associé.

Cependant, les propriétés de la loi normale (voir annexe 6) montrent que  $\frac{Y - y}{u_c(y)}$  suit approximativement une loi normale centrée réduite dès que l'une des conditions suivantes est vérifiée :

- $u_c(y)$  n'est pas dominée par une composante d'incertitude-type obtenue par une évaluation de type A fondée sur quelques observations, ni par une composante d'incertitude-type obtenue par une évaluation de type B fondée sur moins de trois lois rectangulaires ;
- les composantes de  $u_c^2(y)$  fondées sur des lois normales sont significativement beaucoup plus grandes que toute autre composante.

Pour de nombreux mesurages pratiques dans une large étendue de domaines, les conditions suivantes prédominent :

- l'estimation  $y$  du mesurande  $Y$  est obtenue à partir des estimations  $x_i$  d'un nombre significatif de grandeurs d'entrée  $X_i$  qui peuvent être décrites par des lois de probabilités raisonnables telles que des lois normales ou rectangulaires ;
- les incertitudes-types  $u(x_i)$  de ces estimations, qui peuvent être obtenues par des évaluations de type A ou de type B, contribuent de manière comparable à l'incertitude-type composée  $u_c(y)$  du résultat de mesure  $y$  ;
- l'approximation linéaire supposée par la loi de propagation de l'incertitude est convenable.

Dans ces conditions, on peut supposer que la loi de probabilité suivie par la variable  $Y$  est normale en raison du théorème limite central (voir annexe 6) ; et  $u_c(y)$  peut être considéré comme une estimation raisonnablement fiable de l'écart-type de cette loi normale.

On sait que si une variable aléatoire  $X$  suit une loi normale de moyenne  $m$  et d'écart type  $\sigma$ , alors la probabilité que l'on ait  $m - 1,96\sigma \leq X \leq m + 1,96\sigma$  est égale à 0,95.

Dans la pratique, on prend le plus souvent  $k = 2$  ce qui signifie que dans les conditions décrites ci-dessus on peut penser raisonnablement que la procédure suivie a plus de 95% de chances d'aboutir à un intervalle contenant effectivement la « valeur vraie » de  $Y$  ou encore que : l'intervalle  $[y - 2u_c(y); y + 2u_c(y)]$  contient environ 95% des valeurs que l'on peut raisonnablement associer à la grandeur  $Y$ .

On dit alors que l'incertitude élargie  $U = 2 u_c(y)$  définit un intervalle  $[y - U(y); y + U(y)]$  ayant un niveau de confiance d'environ 95 %.

Dans certains cas on peut prendre  $k = 3$  et considérer qu'avec  $U = 3 u_c(y)$  on définit un intervalle ayant un niveau de confiance d'environ 99 %.

Si les conditions d'expérimentation précisées ci-dessus ne sont pas vérifiées, on ne peut pas faire l'hypothèse que la variable  $Y$  suit approximativement une loi normale. Dans ce cas on doit affiner la recherche de la loi de  $Y$  et recourir à d'autres lois comme la loi de Student.

## B. Présentation des résultats

### 1. Arrondissement

Les résultats mesurés ou calculés peuvent comporter des chiffres qui n'ont pas de sens en regard à l'incertitude déterminée sur la mesure.

Il est donc nécessaire de procéder à un arrondissement du résultat obtenu.

Les règles à appliquer sont données ci-dessous :

- Si l'incertitude élargie est  $U$ , alors on arrondit de façon que l'erreur due à l'arrondissement soit inférieure à 1/10 de l'incertitude  $U$  retenue ;
- Le dernier chiffre retenu suit les règles d'arrondissement « au plus près » habituelles.

Remarques :

Pour conserver une cohérence, les incertitudes seront données avec au plus deux chiffres significatifs. Tout arrondissement des incertitudes se fera, par prudence, par excès.

Pour limiter le cumul d'erreurs sur les arrondis, l'arrondissement est effectué sur le résultat final. Pour les calculs intermédiaires on gardera donc des chiffres qui peuvent être non significatifs.

Exemples

Si  $y = 12,3257$  et  $u_c(y)$  estimé à 0,232 avec  $k = 2$ . Alors  $U = 0,464$ . L'erreur d'arrondissement devra alors être inférieure à 0,04. On arrondit donc au 1/100 près. Par conséquent, on prendra :

$$y = 12,32 \pm 0,47$$

Si  $y = 123,385$  et  $u_c(y)$  estimé à 2,892 avec  $k = 2$ . Alors  $U = 5,784$ . L'erreur d'arrondissement devra alors être inférieure à 0,57. On arrondit donc au 1/10 près. Par conséquent, on prendra :

$$y = 123,4 \pm 5,8$$

Cependant, le calcul d'incertitude, basé sur des estimations s'appuie sur des approximations (connaissance des lois) et des connaissances imparfaites (nature des sources d'erreurs). Il est lui-même soumis à incertitude...

On peut estimer que dans le cas d'une estimation d'une grandeur comme la moyenne de  $n$  valeurs, le pourcentage d'incertitude sur l'incertitude proposée est  $r \approx \frac{1}{\sqrt{(2n-1)}}$  (voir annexe 8)

Ainsi l'incertitude sur l'incertitude déterminée sur la moyenne de 10 valeurs est de l'ordre de 25%, et celle sur la moyenne de 50 valeurs de l'ordre de 10%.

Ce résultat justifie le choix qui a été fait dans le document proposé par l'Inspection Générale de Physique de ne conserver qu'un seul chiffre significatif.

On pense souvent que les méthodes d'évaluation de type A de l'incertitude, déterminée à partir de données statistiques donnent des résultats plus fiables que ceux obtenus à partir d'hypothèses sur les lois de probabilités suivies par les erreurs supposées. L'évaluation ci-dessus de l'imprécision de l'incertitude de type A, relativise fortement cette assertion !

## 2. Présentation des résultats

Lorsqu'on exprime le résultat d'un mesurage et son incertitude, il est bon de donner un maximum d'informations sur les conditions d'obtention des résultats annoncés, par exemple :

- Décrire les méthodes utilisées ;
- Donner la liste des composantes de l'incertitude et la manière dont elles ont été évaluées ;
- Présenter l'analyse des résultats de telle façon que ces derniers puissent être réutilisés ;
- Les résultats préciseront, au minimum, le facteur d'élargissement et le niveau de confiance de l'intervalle estimé.

Par exemple :

$y = 5,324 \text{ mm} \pm 0,046 \text{ mm}$  où l'incertitude exprimée est une incertitude élargie d'un facteur  $k=3$ . Cette incertitude définit un intervalle estimé avoir un niveau de confiance proche de 99 %.

## C. En conclusion

Pour déterminer une incertitude sur un mesurage :

- Choisir une méthode de mesurage.
- Modéliser le mesurage :
  - Déterminer les différentes variables qui entrent dans le mesurage.
  - Déterminer les grandeurs d'influence.
  - Déterminer la fonction mathématique qui lie ces variables.
- Déterminer les composantes de l'erreur sur chacune des variables.
  - Lister les composantes aléatoires et les composantes systématiques.
- Déterminer les incertitudes-types sur chacune des variables.
  - Estimations de type A (statistiques) ou de type B (probabilistes)
  - Éventuelle Incertitude – type composée sur chaque variable
- Déterminer l'incertitude – type composée sur le mesurage.
- Déterminer le facteur d'élargissement (associé à la loi supposée représenter la répartition des valeurs prises par le mesurande).
- Donner l'incertitude élargie sur le mesurage avec ses caractéristiques.

## D. Un exemple

Reprenons l'exemple traité de la dimension en fond de rainure. On a les données suivantes :

La pige  $w_1$  est une pige de 8mm. Le certificat d'étalonnage indique que « l'incertitude à deux sigmas » est de  $2\mu\text{m}$ .

La pige  $w_2$  est une pige de 12mm. Le certificat d'étalonnage indique que « l'incertitude à deux sigmas » est de  $2\mu\text{m}$ .

Les mesures de longueurs sont effectuées à l'aide d'un pied à coulisse au 1/100 dont l'erreur de justesse maximale est de  $30\mu\text{m}$ .

On suppose que la salle de mesure est maintenue à une température régulée de  $20^\circ\text{C}$ .

- la méthode de mesure choisie est celle avec deux piges, afin de s'affranchir du contrôle de la mesure de l'angle.
- la modélisation mathématique est donnée par  $d = \frac{L_1 w_2 - L_2 w_1}{w_2 - w_1}$
- déterminons les sources d'incertitude sur les résultats obtenus :
  - incertitudes liées aux manipulations ;
  - incertitude de justesse de l'appareil de mesure ;
  - incertitude sur le diamètre des piges ;
  - incertitude due à la forme de l'objet (en partie prise en compte dans a) ;
  - incertitude dues aux conditions environnementales : stabilité de la température, différence de température entre la pièce et les piges... (négligée, car faible devant a)).
- déterminons les valeurs des coefficients :

$$\text{On a : } c_1 = \frac{\partial d}{\partial L_1} = \frac{w_2}{w_2 - w_1} = 3, \quad c_2 = \frac{\partial d}{\partial L_2} = \frac{-w_1}{w_2 - w_1} = -2,$$

$$c_3 = \frac{\partial d}{\partial w_1} = \frac{-(L_2 - L_1)w_2}{(w_2 - w_1)^2} = -5,121 \text{ et enfin } c_4 = \frac{\partial d}{\partial w_2} = \frac{(L_2 - L_1)w_1}{(w_2 - w_1)^2} = 3,414$$

En prenant  $w_1 = 8$ ,  $w_2 = 12$ , pour valeur de  $L_1$  la moyenne 52,35 des 10 valeurs et pour valeur de  $L_2$  la moyenne 59,178 des 10 valeurs

- déterminons les incertitudes sur les quatre variables :

L'incertitude-type sur  $L_1$  a déjà été calculée et elle a pour valeur 0,0190

L'incertitude-type sur  $L_2$  se calcule de même et elle a pour valeur 0,0192 avec  $u_A = 0,00780$  mm et

$$u_B = \frac{0,030}{\sqrt{3}} \text{ mm (instrument vérifié)}$$

L'incertitude-type sur  $w_1$ , comme sur  $w_2$  est 0,0010 mm (0,0020 mm représente deux écarts-types)

- Résumons les résultats dans un tableau, les données sont exprimées en mm :

Composantes	Sources d'incertitudes	$u(x_i)$	$c_i$	$ c_i u(x_i)$	Poids relatif
$u(L_1)$		0,0190	3	0,0570	68,2%
$s_9$	Observations répétées	0,00816			
$u_j$	Erreur de justesse du PAC	$0,03/\sqrt{3}$			
$u(L_2)$		0,0192	-2	0,0384	31%
$s_9$	Observations répétées	0,00780			
$u_j$	Erreur de justesse du PAC	$0,03/\sqrt{3}$			
$u(w_1)$	Donnée constructeur (2 sigmas)	0,001	-5,121	0,00512	0,55%
$u(w_2)$	Donnée constructeur (2 sigmas)	0,001	3,414	0,00341	0,25%

On a alors  $u_c^2(d) = (0,0570)^2 + (0,0384)^2 + (0,00512)^2 + (0,00341)^2 = 0,00478\dots$

D'où  $u_c(d) = 0,069\dots$

Avec un coefficient d'élargissement  $k = 2$ , on prendra  $U = 0,14$  mm.

Le poids faible des incertitudes associées à des variables rectangulaires nous permet raisonnablement de faire l'hypothèse que la variable  $d$  suit approximativement une loi normale.

- En conclusion :

L'incertitude sur  $d$  élargie d'un facteur  $k = 2$  vaut 0,14 mm ; cette incertitude définit un intervalle estimé avoir un niveau de confiance proche de 95%.

*Remarque 1 :*

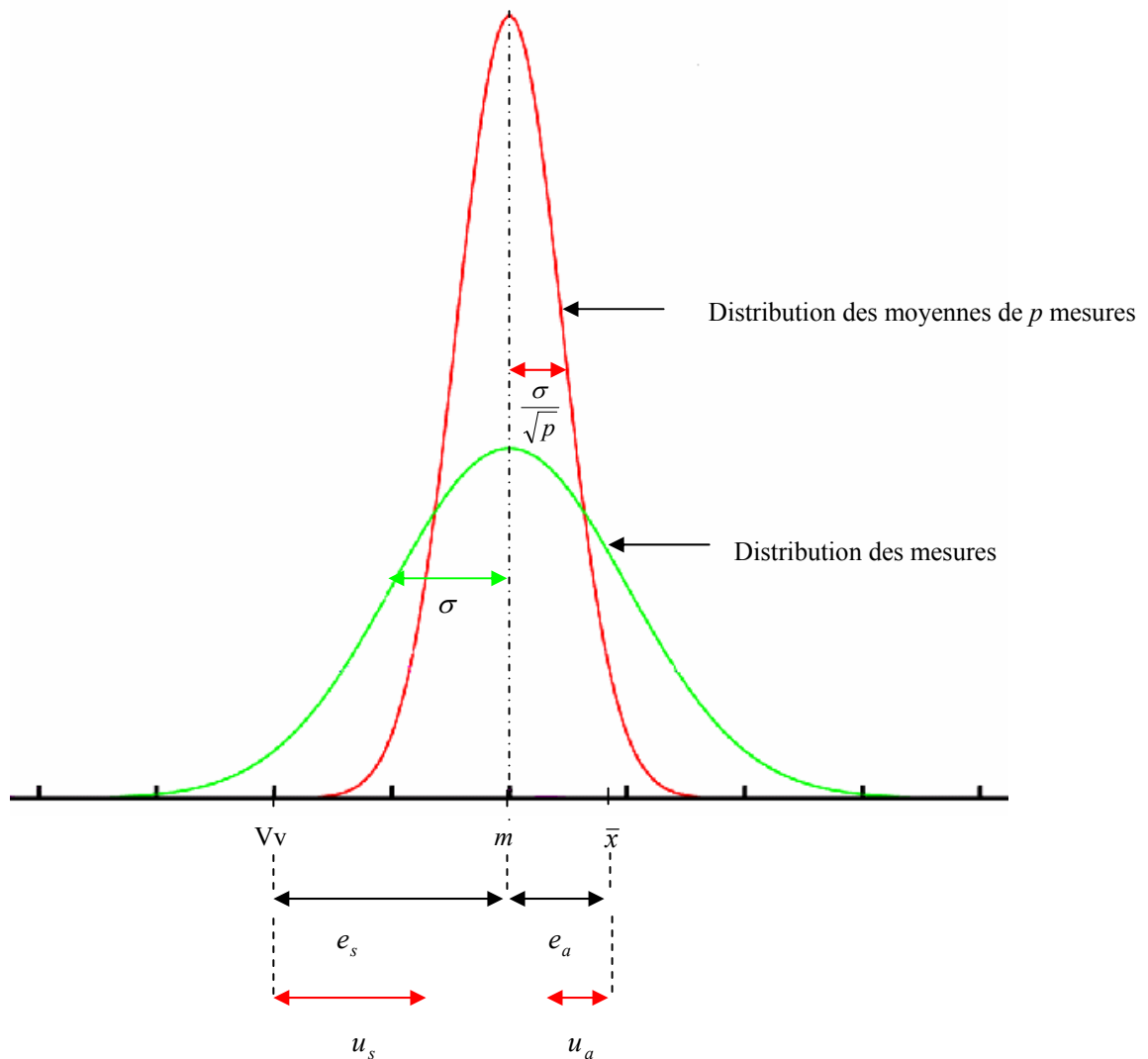
À la lecture de ce tableau, on peut remarquer la part importante de l'incertitude sur les calculs de longueurs ; une analyse de l'incertitude obtenue à partir d'une méthode s'appuyant sur une estimation de l'angle  $\alpha$  permettrait d'évaluer la pertinence du choix de cette méthode.

*Remarque 2 :*

Un logiciel, développé par Jean-Marie Biansan permet d'automatiser les calculs. Il est disponible à l'adresse : <http://jeanmarie.biansan.free.fr/logiciel.html>

### Annexe 1 : Les incertitudes-types sur le mesurage d'une grandeur

#### 1. Mesure sans correction



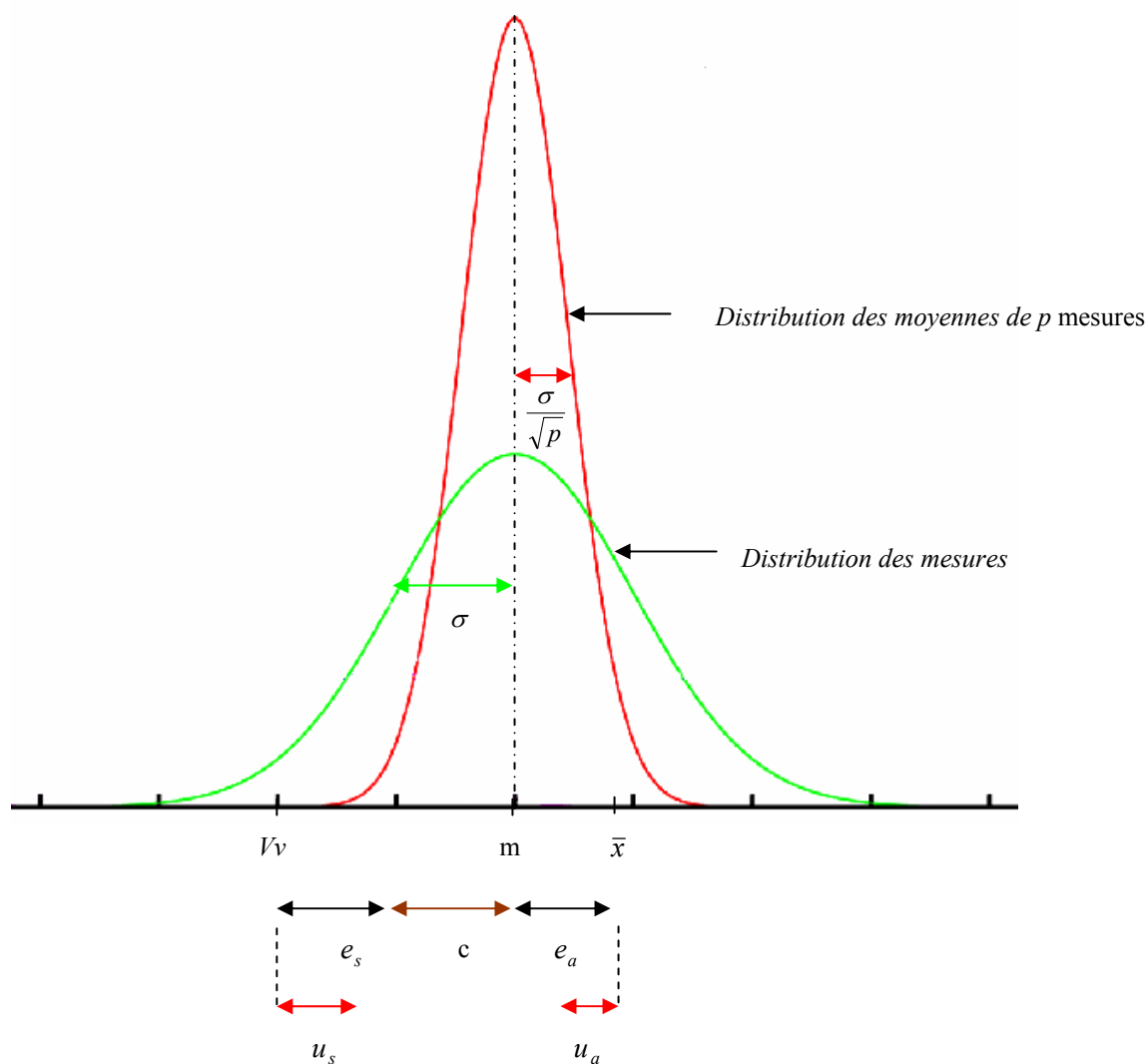
La grandeur est estimée par la moyenne  $\bar{x}$  de  $p$  mesures

- $m$  est la moyenne de la population infinie des mesures
- $\sigma$  est l'écart-type de la population infinie des mesures
- $V_v$  est la valeur vraie (idéale)
- $e_s$  est l'erreur systématique sur le mesurage
- $e_a$  est l'erreur aléatoire sur la moyenne  $\bar{x}$  des  $p$  mesures
- $u_a$  est l'incertitude-type de l'erreur aléatoire sur les  $p$  mesures
- $u_s$  est l'incertitude-type de l'erreur systématique

**La meilleure estimation de  $x$  est  $\bar{x}$**

**L'incertitude type sur la grandeur  $x$  est  $u(x)$  définie par  $u(x)^2 = u_a(x)^2 + u_s(x)^2$**

## 2. Mesure avec correction



La grandeur est déterminée à partir de la moyenne  $\bar{x}$  de  $p$  mesures

- $m$  est la moyenne de la population infinie des mesures
- $\sigma$  est l'écart-type de la population infinie des mesures
- $V.v$  est la valeur vraie (idéale)
- $c$  est la correction effectuée sur les mesures
- $e_s$  est l'erreur systématique sur le mesurage
- $e_a$  est l'erreur aléatoire sur la moyenne  $\bar{x}$  des  $p$  mesures
- $u_a$  est l'incertitude type de l'erreur aléatoire sur les  $p$  mesures
- $u_s$  est l'incertitude type de l'erreur systématique sur la correction

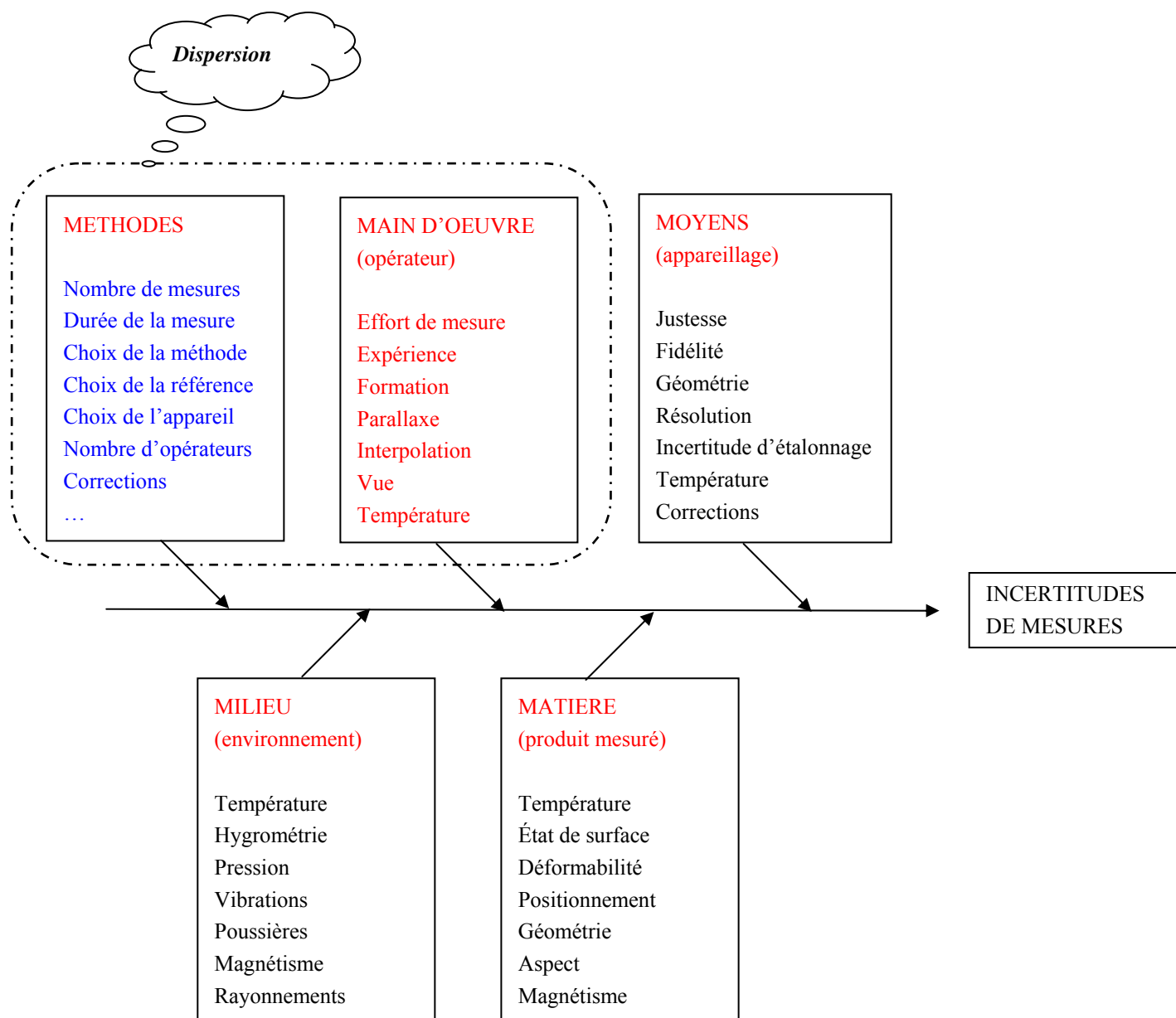
**La meilleure estimation de  $x$  est  $\bar{x} - c$**

**L'incertitude type sur la grandeur  $x$  est  $u(x)$  définie par  $u(x)^2 = u_a(x)^2 + u_s(x)^2$**



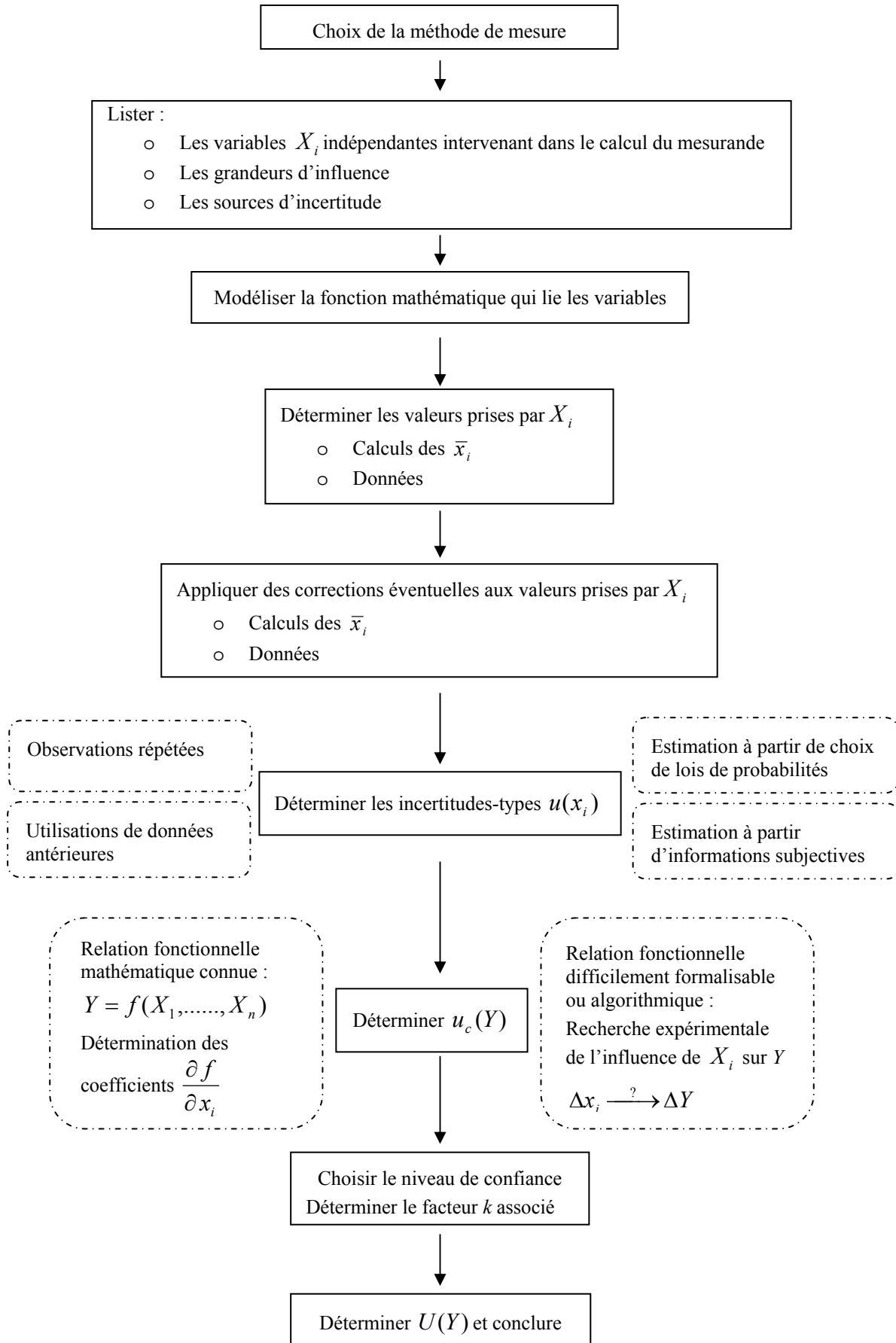
## Annexe 2 : La Démarche de recherche des causes

### Les « cinq M »



Source : Institut Méditerranéen de la Qualité

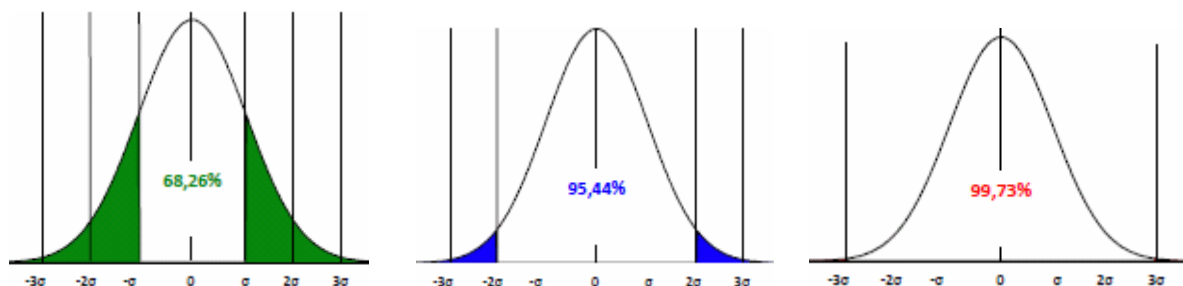
### Annexe 3 : Démarche de détermination d'une incertitude sur une grandeur Y



## Annexe 4 : Un rappel des lois de probabilité

### 1. Lois normales

Ces lois sont d'une grande importance car elles se trouvent être lois limites de la moyenne de variables aléatoires dans le cas de nombreuses lois, lors d'observations répétées de manière indépendante.



allure de la densité de  $X - m$

Si une variable  $X$  suit une loi normale de moyenne  $m$  et d'écart-type  $\sigma$  on a alors :

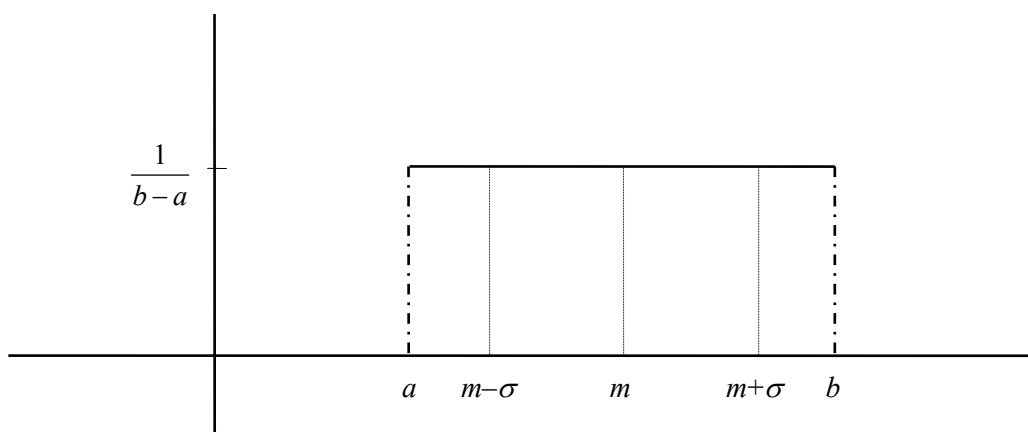
- Probabilité  $(-\sigma \leq X - m \leq \sigma) \approx 0,68$
- Probabilité  $(-2\sigma \leq X - m \leq 2\sigma) \approx 0,95$
- Probabilité  $(-3\sigma \leq X - m \leq 3\sigma) \approx 0,997$

Cette dernière inégalité indique que quasiment la totalité des données sont situées entre  $-3\sigma$  et  $3\sigma$ . On dit parfois que l'étendue des valeurs représente  $6\sigma$ , ce qui permet de donner rapidement une estimation de l'écart-type en divisant cette étendue par 6.

### 2. Lois rectangulaires (ou lois uniformes)

Une variable  $X$  suit une loi uniforme sur un intervalle  $[a, b]$  si sa fonction de densité  $f$  est définie par

$$f(x) = \frac{1}{b-a} \text{ pour } a \leq x \leq b \quad \text{et} \quad f(x) = 0 \text{ sinon}$$



allure de la densité

La moyenne de  $X$  est  $\frac{b+a}{2}$  et son écart-type est  $\sigma = \frac{b-a}{2\sqrt{3}}$

Par exemple dans l'utilisation d'un appareil numérique de résolution 0,01 unité et qui donne 12,46 comme affichage, on fait l'hypothèse que le résultat réel est avec la même probabilité  $\frac{1}{10}$  entre 12,455 et 12,456, entre 12,456 et 12,457,.... ou entre 12,464 et 12,465.

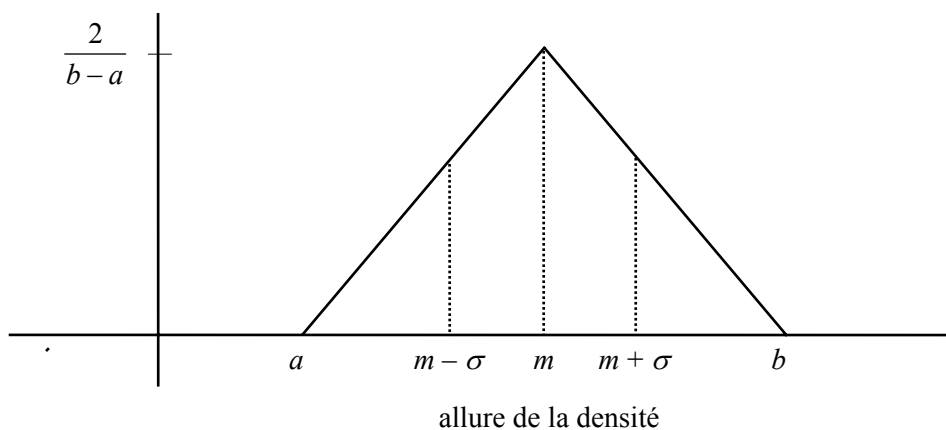
L'incertitude-type liée à la résolution de l'appareil sera  $\frac{0,01}{2\sqrt{3}}$

Une remarque : cette loi est celle du « pire », c'est-à-dire que pour des valeurs que l'on sait comprises entre  $a$  et  $b$ , c'est la loi qui a le plus grand écart-type.

### 3. Lois triangulaires

Une variable  $X$  suit une loi triangulaire sur un intervalle  $[a, b]$  si sa fonction de densité  $f$  est définie par :

$$f(x) = \frac{4(x-a)}{(b-a)^2} \text{ pour } a \leq x \leq \frac{b+a}{2} ; \quad f(x) = \frac{4(b-x)}{(b-a)^2} \text{ pour } \frac{b+a}{2} \leq x \leq b ; \quad f(x) = 0 \text{ sinon}$$



La moyenne de  $X$  est  $\frac{b+a}{2}$  et son écart-type est  $\sigma = \frac{b-a}{2\sqrt{6}}$

## Annexe 5 : Les recommandations de détermination d'incertitude de type B

### 1. Résolution d'une indication numérique.

Si la résolution d'un instrument numérique est  $d$ , la valeur  $X$  du signal mesuré peut se situer « avec une égale probabilité » à n'importe quel endroit de l'intervalle allant de  $X - \frac{d}{2}$  à  $X + \frac{d}{2}$ .

Le signal mesuré suit donc une loi rectangulaire de largeur  $d$

On prendra donc :

$$u(x) = \frac{d}{2\sqrt{3}}$$

Exemple : si un instrument de pesage a un dispositif indicateur dont le dernier chiffre significatif correspond à 1g, l'incertitude-type sur la résolution est  $u = \frac{1}{2\sqrt{3}}$  g soit environ 0,29g

### 2. Hystérésis

L'indication d'un instrument peut varier selon que les lectures se font par valeurs croissantes ou décroissantes. L'opérateur prudent note le sens des lectures successives et fait les corrections nécessaires. Cependant le sens de l'hystérésis n'est pas toujours décelable (oscillations autour d'un point d'équilibre par exemple) et on suppose alors que la loi de probabilité suivie par l'hystérésis est une loi rectangulaire.

*Si l'étendue des lectures possibles dues à cette cause est  $\delta x$ , alors  $u(x) = \frac{\delta x}{2\sqrt{3}}$*

### 3. Arrondissement, calculs à précision limitée

L'arrondissement ou la troncature des nombres qui se produit dans les réductions automatiques de données par les ordinateurs peut aussi être source d'incertitude (problème de soustractions de nombres proches par exemple). Si une simulation de données proches sur les grandeurs d'entrées permet de déceler une variation sur la valeur de sortie, on suppose que cette valeur suit une loi rectangulaire.

*Si  $\delta x$  est la plus petite variation de la grandeur de sortie, alors  $u(x) = \frac{\delta x}{2\sqrt{3}}$*

### 4. Incetitude-type sur un instrument vérifié

Si le métrologue utilise un instrument vérifié, ce dernier est conforme à une classe qui est définie par une limite  $\pm\alpha$ .

Si on suppose que les valeurs affichées suivent une loi rectangulaire, on prendra alors

$$u(x) = \frac{\alpha}{\sqrt{3}}$$

Cependant, s'il y a des raisons de penser que les valeurs situées près de 0 sont plus probables que celles situées près de  $\alpha$  ou  $-\alpha$  on pourra penser que la loi de propagation de l'erreur est normale ou par prudence triangulaire ; on prendra alors :

$$u(x) = \frac{\alpha}{3} \text{ (choix de normalité) ou } u(x) = \frac{\alpha}{\sqrt{6}}$$

## 5. Incertitude-type sur un instrument étalonné

Dans le cas d'un instrument étalonné, le certificat d'étalonnage annonce une incertitude  $U$ . En fait cette valeur  $U$  est égale à l'incertitude-type  $u(x)$  multipliée par un facteur d'élargissement  $k$  (pratique justifiée en fin de document).

On prendra donc

$$u(x) = \frac{U}{k}$$

En principe, le facteur d'élargissement devrait être précisé avec l'incertitude donnée par le constructeur.

Si aucune précision n'est faite, on supposera que  $k = 2$ .

Remarque : on suppose en général que la détermination de l'incertitude-type obtenue s'appuie sur une loi de distribution de l'erreur normale.

## 6. Incertitude-type due aux effets de la température

En métrologie dimensionnelle, la température est un facteur d'influence de premier ordre. Pour en minimiser les effets, une méthode souvent employée est de comparer le mesurande à un étalon de référence de même longueur dont on connaît la valeur « vraie » à la température de 20°C.

La température intervient à différents niveaux :

- écart de température avec 20°C noté  $\Delta t$

La loi de propagation de l'erreur sur des grandeurs sous contrôle (par exemple température dans un bain régulé) demande une analyse fine de la nature des régulations ou une analyse statistique pertinente.

Dans un milieu climatisé, la loi de propagation de l'erreur sur l'écart à 20°C est en général assimilée à une loi en « dérivée d'arcsinus ». Dans ces conditions, on prendra :

$$u(\Delta t) = \frac{\Delta t}{\sqrt{2}}$$

- écart de température entre le mesurande et les cales étalons noté  $\delta t$

En général, on fait l'hypothèse que  $\delta t$  suit une loi normale et on prend alors

$$u(\delta t) = \frac{\delta t}{3}$$

## Annexe 6 : La loi normale

L'établissement, au début du XIXe siècle, de la loi "normale", dont l'usage est fondamental en statistique, s'est fait par deux voies : celle, dans le cadre de la "théorie des erreurs", de la méthode des moindres carrés, qui aboutit avec **Carl Friedrich Gauss** (1777-1855), et celle des théorèmes limites, avec l'énoncé d'une première version du théorème limite central par **Pierre Simon de Laplace** (1749-1827).

L'astronomie et la géodésie sont à l'origine des questions théoriques sur la répartition des erreurs de mesure "accidentelles" (que l'on peut qualifier d'aléatoires), dues à l'addition de nombreux facteurs indépendants (conditions de la mesure, erreurs de lecture, de visée...), qui peuvent induire une erreur dans un sens ou dans l'autre. L'objectif est de pouvoir aller au delà de la précision d'un instrument, en "combinant" plusieurs mesures de la même quantité, de façon à calculer la "meilleure estimation" de cette dernière.

**Adrien-Marie Legendre** (1752-1833) publie en 1805 la méthode consistant à minimiser la somme des carrés des écarts. Indépendamment, **Gauss**, alors directeur de l'observatoire de Göttingen, parvient, dans le cadre de l'étude des orbites planétaires, à cette même **méthode des moindres carrés**, dit-il dès 1794 (il en conteste la paternité à **Legendre**, mais ne publiera qu'en 1809). L'originalité de **Gauss** est d'établir les liens qui existent entre cette méthode et les lois de probabilité, aboutissant à la "loi gaussienne" :

Il raisonne ainsi :

Soit une quantité  $\theta$  inconnue, pour laquelle on possède plusieurs mesures  $x_1, x_2, \dots, x_n$ . On cherche à minimiser la somme des carrés des écarts  $\sum_{i=1}^n (\theta - x_i)^2$ . En considérant cette quantité comme une fonction de  $\theta$  un simple calcul de dérivée montre que la valeur de  $\theta$  rendant minimale la somme des carrés des erreurs est la moyenne :  $\bar{x} = \frac{x_1 + \dots + x_n}{n}$

En envisageant la question d'un point de vue probabiliste, on considérera que les erreurs  $e_1 = x_1 - \theta, \dots, e_n = x_n - \theta$  sont des réalisations de  $n$  variables aléatoires indépendantes  $E_1, \dots, E_n$  de même loi continue de densité  $f$ , dépendant de la valeur inconnue  $\theta$ .

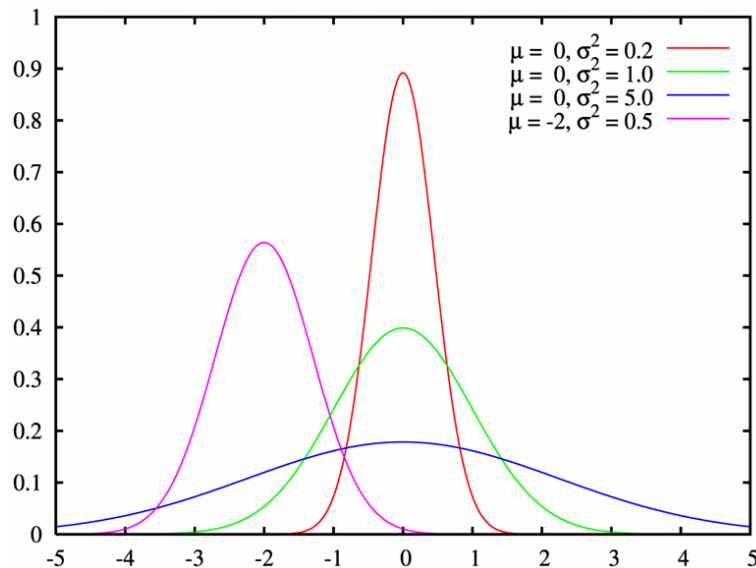
Pour  $\theta$  donné, la probabilité d'effectuer des erreurs à la première mesure entre  $e_1$  et  $e_1 + de_1$  vaut environ  $f(e_1) de_1$ , et en vertu de l'indépendance des mesures, la probabilité que les erreurs se situent entre  $e_1$  et  $e_1 + de_1, \dots, e_n$  et  $e_n + de_n$  est alors,  $f(e_1)f(e_2) \dots f(e_n) de_1 de_2 de_n$ .

On peut alors retourner le raisonnement (à la façon de *Bayes*) et se demander, les mesures  $x_1, \dots, x_n$  étant connues, quelle est la valeur de  $\theta$  la plus vraisemblable. C'est à dire, quelle est la valeur de  $\theta$  qui rendra maximale la probabilité d'observation des mesures  $x_1, \dots, x_n$  (réellement observées) donc des erreurs  $e_1, \dots, e_n$ . Il s'agit de rechercher  $\theta$ , donc  $f$ , de sorte que  $f(e_1)f(e_2) \dots f(e_n) de_1 de_2 de_n$  soit maximum (cette démarche est nommée "**maximum de vraisemblance**").

Sachant que la moyenne arithmétique  $\bar{x} = \frac{x_1 + \dots + x_n}{n}$  correspond à la valeur recherchée de  $\theta$ , **Gauss** en déduit par un calcul algébrique assez simple que la fonction  $f$  est de la forme  $f(x) = Ce^{-kx^2}$ .

La fonction  $f$  étant une fonction de densité, on montre que  $C = \frac{1}{\sqrt{2\pi}}$  et plus généralement, on appelle

loi normale de moyenne  $m$  et d'écart-type  $\sigma$ , la loi de densité  $f$  définie par  $f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-m}{\sigma}\right)^2}$



Exemples de densité de lois normales de moyenne  $\mu$  et de variance  $\sigma^2$

**Ainsi lorsque, lors de mesures, l'addition de plusieurs facteurs aléatoires indépendants (et sensiblement équivalents) induit des erreurs, celles-ci se répartissent selon la loi de Gauss et la moyenne arithmétique des mesures fournit l'estimation qui minimise la somme des carrés des erreurs.**

L'approche de **Laplace** se situe dans la voie des **lois limites**. Il montre, que sa "*seconde loi des erreurs*" (qui est la loi de Gauss présentée précédemment) approche la distribution des moyennes arithmétiques de  $n$  erreurs indépendantes de même loi.

*Laplace* et *Gauss* réalisent ainsi, au début du XIX<sup>ème</sup> siècle, une synthèse entre l'approche empirique des moindres carrés et celle, probabiliste, des lois limites. Avec *Laplace*, la loi normale s'impose comme presque universelle. En effet, même si la distribution individuelle des erreurs ne suit pas une loi normale, celle des moyennes des erreurs suit approximativement, sous certaines conditions (indépendance, lois identiques), une loi normale.

Plus précisément on peut énoncer un résultat qui est central dans la théorie des probabilités :

Soient  $X_i$  des variables aléatoires indépendantes de même moyenne  $m$  et d'écart type  $\sigma$ . Pour  $n$  suffisamment grand, la variable  $\bar{X} = \frac{1}{n} \sum_{i=1}^n X_i$  suit approximativement la loi normale de même moyenne  $m$  et d'écart-type  $\frac{\sigma}{\sqrt{n}}$ .

Ce qui signifie que si les  $X_i$  suivent une loi **quelconque** de moyenne  $m$  et d'écart-type  $\sigma$  alors la variable aléatoire  $\bar{X} = \frac{1}{n} \sum_{i=1}^n X_i$  suit approximativement une loi normale.

Ainsi, on peut schématiquement dire que si on considère que l'ensemble des mesures d'une grandeur est associé à une distribution par exemple uniforme, alors si à chaque série de mesures on associe la moyenne de ces mesures, la distribution de l'ensemble de ces moyennes est approximativement normale.



Deux remarques :

- Si la loi suivie par les variables  $X_i$  est normale, la loi suivie par  $\bar{X} = \frac{1}{n} \sum_{i=1}^n X_i$  est exactement normale.
- plus la valeur de  $n$  est grande, meilleure est l'approximation.

En fait ce résultat se déduit d'un théorème plus général de convergence très utilisé dans la théorie probabiliste des erreurs, le théorème limite central (TLC) :

Soient  $X_i$  des variables aléatoires indépendantes de même moyenne  $m$  et d'écart type  $\sigma$ , on note

$$Z_n = \frac{\sum_{i=1}^n X_i - nm}{\sigma\sqrt{n}}$$

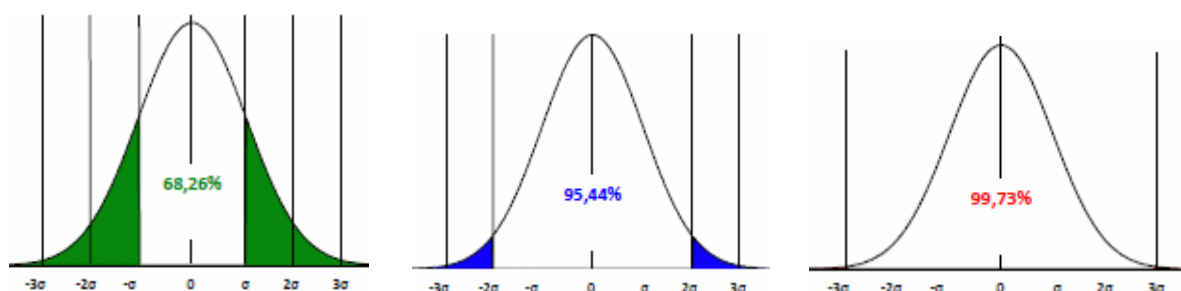
et  $T$  une variable qui suit la loi normale de moyenne nulle et d'écart type 1.

Alors pour toute valeur de  $z$ ,  $P(Z_n < z)$  tend vers  $P(T < z)$  quand  $n$  tend vers  $+\infty$ .

On dit que la variable  $Z_n = \frac{\sum_{i=1}^n X_i - nm}{\sigma\sqrt{n}} = \frac{\frac{\sum_{i=1}^n X_i}{n} - m}{\frac{\sigma}{\sqrt{n}}}$  converge en loi vers la variable  $T$ .

C'est ce théorème qui permet de justifier ce que l'expérimentation permet de constater : dans des conditions de mesurages où les incertitudes prennent des valeurs du même ordre, alors la variable aléatoire qui modélise l'erreur suit approximativement une loi normale.

### Propriétés de la loi normale



Si une variable  $X$  suit une loi normale de moyenne  $m$  et d'écart-type  $\sigma$  alors on a alors :

- Probabilité  $(-\sigma \leq X - m \leq \sigma) \approx 0,68$
- Probabilité  $(-2\sigma \leq X - m \leq 2\sigma) \approx 0,95$
- Probabilité  $(-3\sigma \leq X - m \leq 3\sigma) \approx 0,997$

Cette dernière inégalité indique que quasiment la totalité des données sont situées entre  $-3\sigma$  et  $3\sigma$

On dit parfois que l'étendue des valeurs représente  $6\sigma$ , ce qui permet de donner une estimation grossière de l'écart-type en divisant cette étendue par 6.

Il va être important de pouvoir vérifier si la loi considérée est normale, c'est ce qui est fait au travers de « tests de normalité ».

## Tests de normalité

Dans un test de normalité on compare la distribution d'un échantillon de données avec la distribution « théorique » attendue sous l'hypothèse que cet échantillon correspond à un tirage de données des valeurs prises par une variable aléatoire qui suit une loi normale.

Le plus simple est graphique (droite de Henry), ce qui a pour intérêt une grande simplicité mais pour l'inconvénient de mal maîtriser le niveau d'approximation. D'autres tests peuvent être mis en œuvre comme le test du Khi2 qui permet de se donner une règle de décision ou encore celui de Kolmogorov qui affine le risque d'erreur.

On ne développera ci-dessous que le test de Henry.

Le principe de la droite de Henry (du nom d'un artilleur qui avait mis au point cette méthode pour l'étude de précision des tirs) est simple ; l'idée étant, à l'aide d'un changement de variable, d'ajuster les données avec une droite, et d'utiliser la régression linéaire selon les moindres carrés pour quantifier la qualité de l'ajustement d'une distribution statistique observée avec celle d'une loi normale.

Pour cela on compare la fréquence cumulée des données jusqu'à une valeur  $x_i$  avec la probabilité « théorique » jusqu'à cette même valeur.

Ainsi, si on considère une variable aléatoire  $X$  suivant la loi  $\mathcal{N}(m; \sigma)$ , sa fonction de répartition  $F$  est donnée, pour tout  $x \in \mathbb{R}$ , par :

$$y = F(x) = P(X \leq x) = P\left(\frac{X - m}{\sigma} \leq \frac{x - m}{\sigma}\right) = \Pi(t) \text{ avec } t = \frac{x - m}{\sigma} \text{ et } \Pi \text{ la fonction de répartition de la loi normale centrée réduite dont les valeurs sont données dans des tables.}$$

Si on considère maintenant la série des données, l'analogue de la fonction de répartition est la fréquence cumulée ; pour une valeur  $x_i$  de la distribution statistique, on note  $y_i$  la fréquence cumulée croissante.

Si la loi de la distribution des données est normale, les fréquences  $y_i$  cumulées des données jusqu'à une valeur  $x_i$  doivent correspondre à des probabilités données par la table de la loi normale.

C'est-à-dire que si on note  $t_i$  la valeur, donnée par lecture inverse de la table de la loi normale centrée réduite pour la fréquence cumulée  $y_i$  (donc telle que  $y_i = \Pi(t_i)$ ), alors le nuage de points  $(x_i ; t_i)$

devrait être aligné ou tout du moins correctement ajusté par la droite d'équation  $t = \frac{x - m}{\sigma}$  que l'on nomme **droite de Henry**.

Remarque : le tracé de cette droite permet également de donner une estimation de  $m$  et  $\sigma$ .

Prenons un exemple :

Un mesurage fournit les 50 données suivantes :

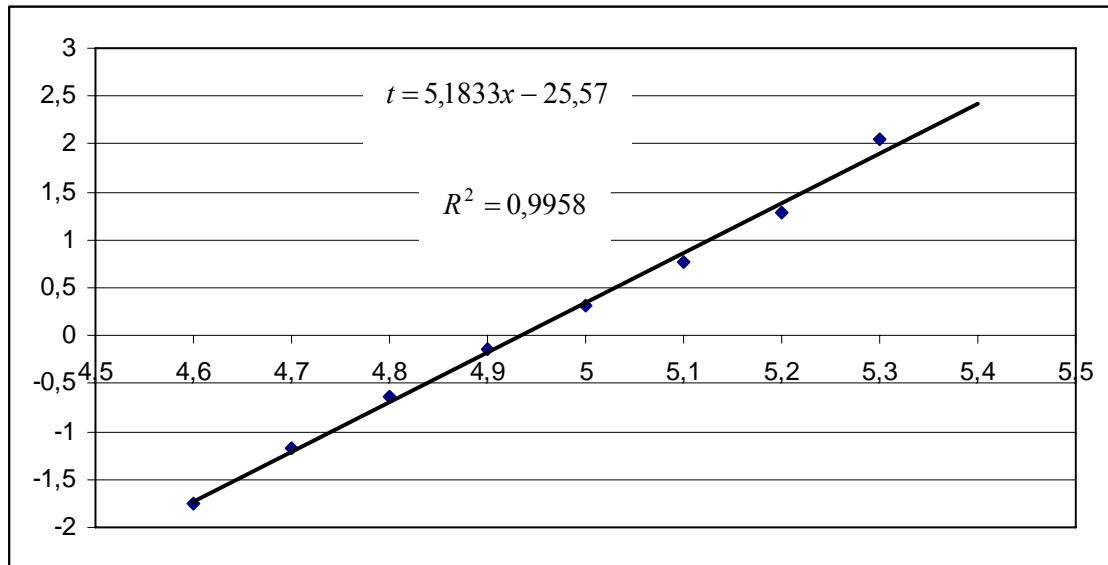
$x_i$	4,6	4,7	4,8	4,9	5	5,1	5,2	5,3	5,4
$n_i$	2	4	7	9	9	8	6	4	1

On s'interroge sur la normalité des données.

On complète le tableau par les fréquences, les fréquences cumulées et les valeurs de  $t_i$  associées, lues dans une table donnant la fonction de répartition de la loi normale centrée réduite.

$x_i$	4,6	4,7	4,8	4,9	5	5,1	5,2	5,3	5,4
$n_i$	2	4	7	9	9	8	6	4	1
Fréquences $f_i$	0,04	0,08	0,14	0,18	0,18	0,16	0,12	0,08	0,02
Fréquences cumulées $y_i$	0,04	0,12	0,26	0,44	0,62	0,78	0,90	0,98	1
$t_i$	-1,75	-1,17	-0,64	-0,15	0,31	0,77	1,28	2,05	

Il suffit alors de tracer le nuage des points  $(x_i ; t_i)$  et de déterminer l'ajustement linéaire par la méthode des moindres carrés, par exemple avec l'aide d'un tableur. La valeur du coefficient de corrélation est un indicateur de la qualité de la « normalité ».



Le nuage de points  $(x_i ; t_i)$  est approximativement « linéaire » et le coefficient de corrélation est très proche de 1. On peut accepter le fait que les données sont issues d'une distribution approximativement normale.

On a  $\frac{1}{\sigma} = 5,1833$  et  $\frac{m}{\sigma} = 25,57$  d'où on déduit que l'on peut estimer  $\sigma$  par 0,19 et  $m$  par 4,93.

Remarque :

Il existe un papier dit « gaussio-arithmétique » qui permet de contrôler « au jugé » la normalité des données en plaçant directement sur un graphique, non pas les points  $(x_i ; t_i)$  mais les points  $(x_i ; y_i)$ . Sa grande simplicité fait son intérêt et il est encore employé.

## Annexe 7 : Incertitude composée

On note  $\mu_i$  l'espérance  $E(X_i)$  de chacune des variables et  $\mu = (\mu_1, \mu_2, \dots, \mu_n)$ ; en général ces espérances ne sont pas connues mais estimées par  $x_i$  (qui peut être une moyenne ou non d'observations). Si on note  $x = (x_1, x_2, \dots, x_n)$ , on peut estimer  $E(Y)$  par  $f(x)$  ou par la moyenne  $\bar{y}$  de  $n$  observations répétées.

Commençons par le cas d'une fonction linéaire de deux variables  $Y = f(X_1, X_2) = k_1 X_1 + k_2 X_2$ ;

$$Y = k_1 \mu_1 + k_2 \mu_2 + k_1 (X_1 - \mu_1) + k_2 (X_2 - \mu_2) \text{ et } \bar{Y} = k_1 \mu_1 + k_2 \mu_2 = f(\mu)$$

$$\text{D'où } Y = f(\mu) + \frac{\partial f}{\partial X_1}(\mu)(X_1 - \mu_1) + \frac{\partial f}{\partial X_2}(\mu)(X_2 - \mu_2)$$

Ainsi, si on généralise au cas d'une **fonction linéaire** de  $n$  variables,  $Y = k_1 X_1 + k_2 X_2 + \dots + k_n X_n$

Chacun des coefficients  $k_i$  est la dérivée partielle de  $f$  par rapport à la variable  $X_i$ , les dérivées seconde sont toutes nulles et un développement de Taylor de  $f$  au voisinage de  $\mu$  se réduit à :

$$Y = f(\mu) + \sum_{i=1}^n \frac{\partial f}{\partial X_i}(\mu)(X_i - \mu_i) \text{ et par conséquent,}$$

$$(Y - f(\mu))^2 = \left( \sum_{i=1}^n \frac{\partial f}{\partial X_i}(\mu)(X_i - \mu_i) \right)^2$$

$$(Y - f(\mu))^2 = \sum_{i=1}^n \left( \frac{\partial f}{\partial X_i}(\mu) \right)^2 (X_i - \mu_i)^2 + \sum_{i,j=1, i \neq j}^n \frac{\partial f}{\partial X_i}(\mu) \frac{\partial f}{\partial X_j}(\mu) (X_i - \mu_i)(X_j - \mu_j)$$

$$E(Y - \bar{Y})^2 = \sum_{i=1}^n \left( \frac{\partial f}{\partial X_i}(\mu) \right)^2 E(X_i - \mu_i)^2 + \sum_{i,j=1, i \neq j}^n \frac{\partial f}{\partial X_i}(\mu) \frac{\partial f}{\partial X_j}(\mu) E((X_i - \mu_i)(X_j - \mu_j))$$

Soit :

$$\sigma^2(Y) = \sum_{i=1}^n \left( \frac{\partial f}{\partial X_i}(\mu) \right)^2 \sigma^2(X_i) + \sum_{i,j=1, i \neq j}^n \frac{\partial f}{\partial X_i}(\mu) \frac{\partial f}{\partial X_j}(\mu) \text{cov}(X_i, X_j)$$

Dans le cas où les variables  $X_1, X_2, \dots, X_n$  sont indépendantes et  $f$  est linéaire, on a donc

$$\sigma^2(Y) = \sum_{i=1}^n \left( \frac{\partial f}{\partial X_i}(\mu) \right)^2 \sigma^2(X_i) \text{ et par conséquent } u_c^2(y) = \sum_{i=1}^n \left( \frac{\partial f}{\partial X_i}(\mu) \right)^2 u^2(x_i)$$

*Remarque* : dans la pratique on estime  $\mu$  par  $x = (x_1, x_2, \dots, x_n)$  et on remplace la notation  $\frac{\partial f}{\partial X_i}(\mu)$

par  $\frac{\partial f}{\partial x_i}$ . On écrit alors  $u_c^2(y) = \sum_{i=1}^n \left( \frac{\partial f}{\partial x_i} \right)^2 u^2(x_i)$ .

**Dans le cas général**, les erreurs étant considérées comme « petites » devant les variables, on procède à un développement de Taylor au rang 1 de  $f$  au voisinage de  $x = (x_1, x_2, \dots, x_n)$ , ce qui revient à dire qu'en un point, on assimile une courbe à sa tangente, une surface à son plan tangent. On calcule alors l'incertitude composée sur  $y$  à l'aide de la formule présentée ci-dessus.

## Annexe 8 : Incertitude sur l'incertitude

Si nous reprenons le cas de l'estimation de type A d'une grandeur  $X$ , considérée comme une variable aléatoire. On a noté  $(X_1, X_2, \dots, X_n)$  un  $n$ -échantillon de  $X$ , où  $X_i$  représente la variable aléatoire associée à la  $i^{\text{ème}}$  mesure de la grandeur  $X$ ; la variable  $\bar{X} = \frac{1}{n} \sum_{i=1}^n X_i$  est un estimateur de la moyenne de  $X$  et  $S^2 = \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})^2$  est un estimateur de la variance de  $X$ .

On sait alors que  $\frac{S}{\sqrt{n}}$  est un estimateur de l'écart-type de  $\bar{X}$ , qui caractérise l'incertitude sur la grandeur  $X$ .

Par conséquent, pour estimer ce que représente en pourcentage l'incertitude sur cette incertitude sur  $X$ , on détermine le rapport de l'écart type de  $\frac{S}{\sqrt{n}}$  par la moyenne de  $\frac{S}{\sqrt{n}}$ , c'est-à-dire  $r = \frac{\sqrt{\text{var}(\frac{S}{\sqrt{n}})}}{E(\frac{S}{\sqrt{n}})}$

Dans un cas général, ce rapport est difficilement calculable. Cependant, si la variable  $X$  suit une loi normale alors on montre que la variable  $\frac{(n-1)S^2}{\sigma^2}$  suit une loi du  $\chi^2$  à  $(n-1)$  degrés de liberté, et on montre alors que  $r \approx \frac{1}{\sqrt{2(n-1)}}$

(On en trouve une démonstration calculatoire mais originale dans le livre de John Taylor)

Ainsi l'incertitude sur l'incertitude sur la moyenne de 10 valeurs est de l'ordre de 25%, et celle sur la moyenne de 50 valeurs de l'ordre de 10%.

Ce résultat justifie le choix qui a été fait dans le document proposé par l'Inspection Générale de Physique de ne conserver qu'un seul chiffre significatif.

On pense souvent que les méthodes d'évaluation de type A de l'incertitude, déterminée à partir de données statistiques donnent des résultats plus fiables que ceux obtenus à partir d'hypothèses sur les lois de probabilités suivies par les erreurs supposées. L'évaluation ci-dessus de l'imprécision de l'incertitude de type A, relativise fortement cette assertion !

## V. Bibliographie

---

*Guide pour l'expression de l'incertitude de mesure* (GUM version originale de 1995, revue et corrigée en 2008), téléchargeable sur le site du Bureau International des Poids et Mesures (BIPM) à l'adresse <http://www.bipm.org/fr/publications/guides/gum.html>

*Nombres, mesures et incertitudes* (Inspection générale Sciences Physiques et Chimiques). Eduscol [http://media.eduscol.education.fr/file/PC/66/3/Ressources\\_PC\\_nombres\\_mesures\\_incertitudes\\_14466\\_3.pdf](http://media.eduscol.education.fr/file/PC/66/3/Ressources_PC_nombres_mesures_incertitudes_14466_3.pdf)

*Quantifier l'incertitude dans les mesures analytiques*. CITAC, téléchargeable sur le site du Laboratoire National de Métrologie et d'Essais (LNE) à l'adresse : [http://www.lne.fr/publications/eurachem\\_guide\\_incertitude\\_fr.pdf](http://www.lne.fr/publications/eurachem_guide_incertitude_fr.pdf)

DUTARTE Philippe, PIEDNOIR Jean-Louis ; *Enseigner la statistique au lycée : des enjeux aux méthodes*. IREM Université Paris 13

LECOLLINET Michel *Evaluation et expression des incertitudes de mesures*. (CNAM)

MOREAU René. *Mesures, erreurs et incertitudes en physique chimie*. La pluridisciplinarité dans les enseignements scientifiques. Actes du séminaire d'été 9 au 13 juillet 2001

PRIEL Marc, PERRUCHET Christophe. *Estimer l'incertitude, mesures, essais*. AFNOR

ROBERT Claudine, TREINER Jacques. *Incertitude des mesures de grandeur* ; téléchargeable à l'adresse : [http://www.statistix.fr/IMG/pdf/erreurs\\_et\\_dispersion.pdf](http://www.statistix.fr/IMG/pdf/erreurs_et_dispersion.pdf)

TAYLOR John. *Incertitudes et analyse des erreurs dans les mesures physiques*. Dunod 2000

(remarque : ce livre américain n'est pas totalement dans l'approche probabiliste, il est en partie seulement conforme à la norme AFNOR)