EE206 Systèmes de mesure

Module 1: Approche statistique de la mesure

Dr J.-M. Fürbringer

Faculté des Sciences de base

École Polytechnique Fédérale de Lausanne

March 2, 2020

1. Le processus de mesure

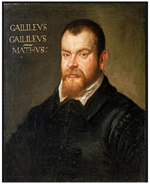
1.1 Acquis d'apprentissage du chapitre

A la fin de cette leçon, vous devez être capable d'expliquer

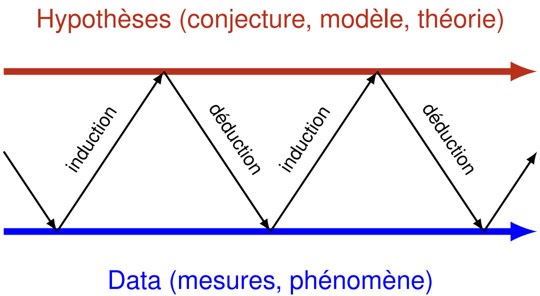
1. De comprendre les concepts statistiques nécessaires à une gestion scientifique d'un processus de mesure
2. De réaliser une analyse visuelle de données issues d'une mesure
3. De réaliser les calculs d'intervalles de confiance pour des résultats expérimentaux primaires et secondaires

1.2 Origine de la science expérimentale

- Invention de la science expérimentale par Galilée (1564-1642) à la Renaissance.
- Un long processus avec plusieurs acteurs, certains célèbres comme Huygens, Ticcho Brahé, Hooke, d'autres oubliés, ou partiellement oubliés comme Rheticus ou l'abbé Nolet
- Galilée affirme que les vérités de la science doivent se confirmer expérimentalement. Les syllogismes de la science antique ont des prémices incomplètes et ne permettent pas des conclusions sûres.
- Par exemple, la théorie des corps pesants et des corps légers d'Aristote ne tient pas compte de la résistance de l'air.
- Aboutis à la théorie du phénomène et du noumène développé par Emmanuel Kant (1724-1804) au XVIIIème.

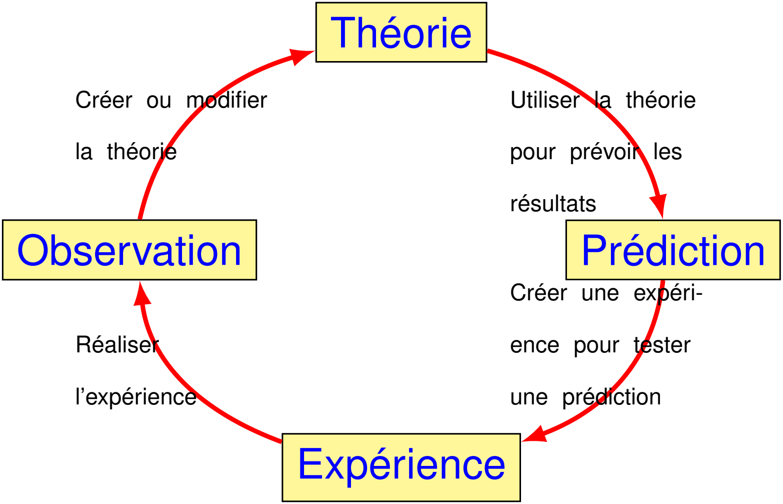


1.3 Expérience vs théorie



from E. G. P. Box, *Statistics for experimenters*, Wiley, 2nd ed., 2005

1.4 La méthode scientifique



1.5 Le rasoir de Ockham [1]

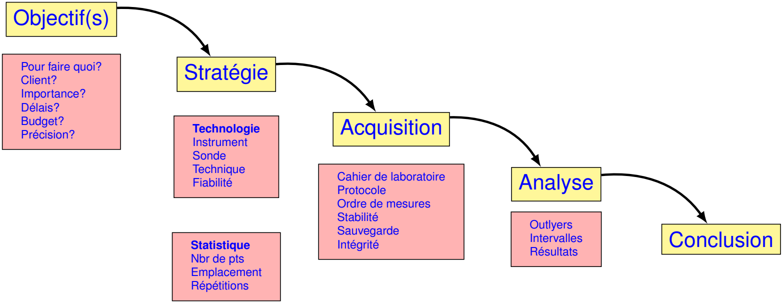
- **De deux modèles prédisant le même résultat, on va choisir le plus simple**
- Approche empirique

[1] Guillaune d'Ockham (1285-1347)

1.6 Sciences et technologies

- Le développement des sciences a permis le développement des techniques
- La disponibilité de grandes sources d'énergie avec le charbon, le pétrole, l'hydroélectrique a amené l'industrialisation
- Ces éléments sont les piliers de l'ère technologique
- Dans l'environnement technologique, la mesure est un élément essentiel tant pour le contrôle (assurer le fonctionnement) que pour le recherche (développer de nouvelles techniques)

1.7 La mesure comme processus



1.8 Trois caractéristiques du monde

- **Multivariable** : les phénomènes sont influencés par plusieurs facteurs.
- **Bruyant** (noisy) : les instruments de mesure ont un niveau de précision fini et des influences non souhaitées interviennent.
- **Avec des interactions** : lorsque l'effet d'un facteur dépend de l'état d'un autre facteur

$\Rightarrow$ Une approche statistique est nécessaire

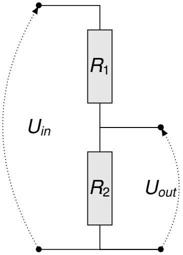
2. Les concepts statistiques

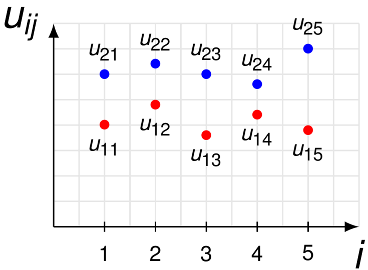
2.1 Variable aléatoire

- La **variable aléatoire** est une concept de base de la statistique que nous allons utiliser pour *modéliser* le résultat d'une mesure
- Il faut distinguer la variable aléatoire à laquelle on a pas accès direct et son échantillonnage qui est le résultat de la mesure
- La mesure de la variable aléatoire X va fournir une série de nombres x_i

2.2 Exemple

- Les tensions U_1 et U_2 par rapport à une référence commune, (une terre) de l'entrée et de la sortie d'un quadripôle sont deux variables aléatoires que l'on peut mesurer N fois et ainsi obtenir N échantillons $\{u_{11}, u_{12}, \dots, u_{1N}\}$ et $\{u_{211}, u_{22}, \dots, u_{2N}\}$ de ces deux variables





2.3 Intput-Output

sur MATLAB

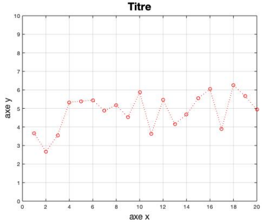
- *X=0:1:20* crée le vecteur $X = [0, 1, 2, \ldots, 20]$
- *X=[. . .]* crée un vecteur avec des valeurs spécifiques
- *T = table(var1, ..., varN, Name, Value)* crée une table
- *T = readtable(FileName.xls','Sheet',SheetName)* charge une table depuis un fichier Excel

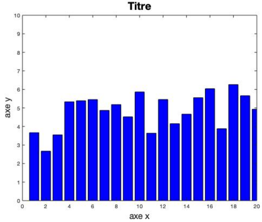
2.0.4 Analyse visuelle

Détecter des motifs, des aberrations, des tendances, etc

Matlab

- *plot(x, y)*
  plot(x, y, LineSpec)
  plot(x_1, y_1, ..., x_n, y_n)
  plot(...,Name,Value)
- *bar(x, y)*
  bar(..., Width)
  bar(..., Style)
  bar(..., Name, Value)
- le style *LineSpec* est constitué de trois caractères pour définir la ligne, le symbole et la couleur, par exemple ':or' correspond à une ligne en pointillé,un symbole en forme de petit cercle, le tout en rouge





2.5 Classer les données

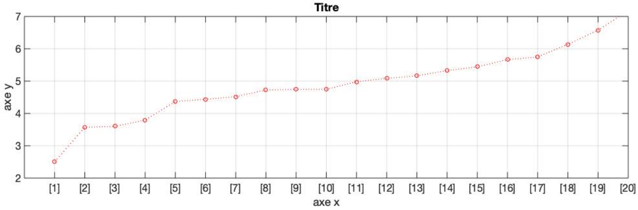
$$x_1 \; x_2 \; \ldots x_N \quad \Rightarrow \quad x_{(1)} < x_{(2)} < \cdots < x_{(N)}$$

MATLAB

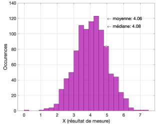
- *[B,Index]=sort(A,dim,direction)*
- *[B,Index]=sortrows(A,col,direction)*
- *[tblB,Index]=sortrows(tblA,col,direction)*

  A, B : matrice
  tblA, tblB : table
  dim : la direction pour effectuer le classement (1,2, . . .)
  col : la colonne de référence pour classer
  direction: 'ascend' or 'descend'

2.0.6 Plot des données classées



2.7 Histogramme des données

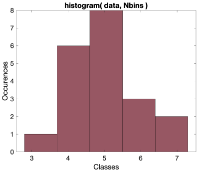


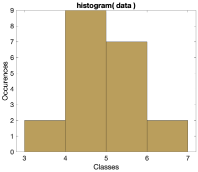
- Représentation graphique permettant de représenter la répartition d'une variable en la représentant par des colonnes verticales les occurrences dans des intervalles ou des catégories qu'on appelle des *classes*.
- Le nombre de valeurs relevées doit être suffisant: plus le nombre est élevé, plus l'interprétation sera aisée.
- Il existe plusieurs manières de déterminer le nombre optimal de classe, K, la plus simple étant $K \approx \sqrt(N)$
- L'amplitude, w est la différence entre la valeur maximale et la valeur minimale
- L'amplitude de chaque classe n'est pas obligatoirement la même

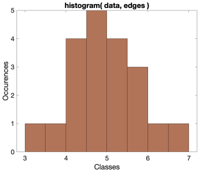
2.8 Exemple d'histogrammes

MATLAB

- *histogram(data,'FaceColor','# EDB120')*
- *histogram(data,5,'FaceColor','# A2142F')*
- *histogram(data,3:.5:7,'Normalization','probability'),...*







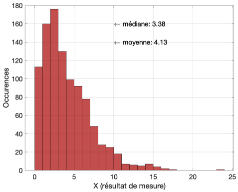
2.9 La moyenne et la médiane

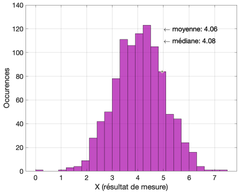
- La **moyenne** une mesure de la position d'une variable aléatoire dans son espace des valeurs

$$\bar{x} = \frac{1}{N} \sum_{i=1}^{N} x_i \tag{2.1}$$

avec N le nombre de mesures et x_i le i-ème résultat

- La **médiane** représente aussi la position de la variable aléatoire mais en séparant son domaine en deux intervalles cumulant chacun 50 % des résultats de mesure: $x_{\left[\frac{N}{2}\right]}$ si N est pair, sinon $\frac{1}{2}\left(x_{\left[\frac{N-1}{2}\right]} + x_{\left[\frac{N+1}{2}\right]}\right)$
- Lorsque la distribution est symétrique, la médiane est égale à la moyenne





2.10 La variance et l'écart-type

- $var(X)$ est une mesure de la dispersion de la variable X
- s'écrit aussi S_x^2
- L'unité de la variance est le carré de celle de la variable

$$S_x^2 = \frac{1}{N-1}\sum_{i=1}^{N}(x_i - \bar{x})^2 \tag{2.2}$$

- L'écart-type S_x est la racine de la variance

$$S_x = \sqrt{\frac{1}{N-1}\sum_{i=1}^{N}(x_i - \bar{x})^2} \tag{2.3}$$

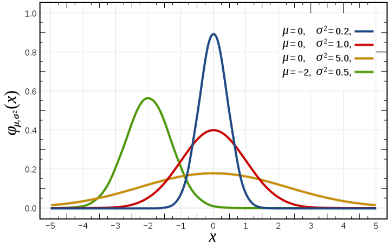
2.11 Quelques fonctions sur MATLAB

- Médiane (*median*)
- Moyenne (*average*)
- Plage (*range*)
- Eccart-type (*standard deviation*)
- Variance (*variance*)
- Quartile (*quartile*)
- Centile (*percentile*)

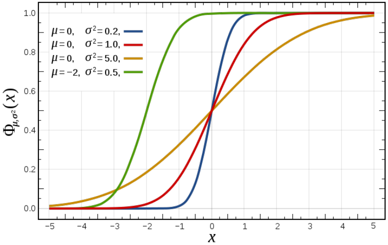
MATLAB

M = median(A)
M = mean(A)
R = range(A)
S = std(A)
v = var(A)
Y = quantile(X,p)

2.12 La distribution Normale N(μ, σ)

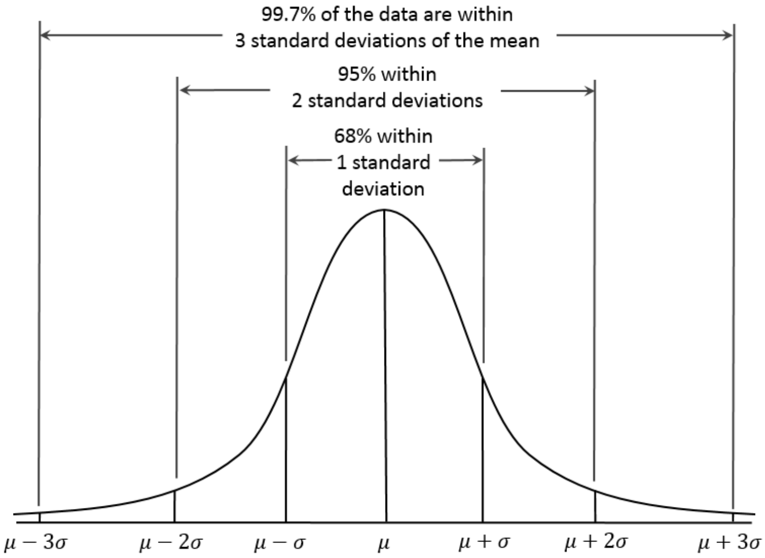


loi de Gauss-Laplace



$$f(x) = \frac{1}{\sqrt{\;}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2} \tag{2.4}$$

2.12 La distribution Normale N(μ, σ)

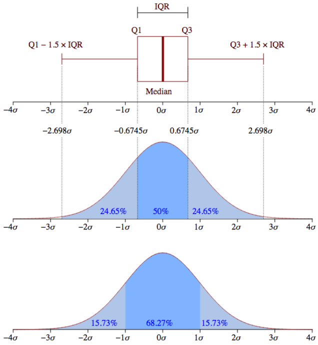


2.12 La distribution Normale

MATLAB

- Génération de nombre aléatoire (utile par exemple pour simuler un processus de mesure)
  X=randn(Ni,Nj)
- La densité de probabilité (probability density function)de la loi Normale aux points *x*
  p=pdf('Normal',x,mu,sigma)
- La fonction de répartition (cumulative density function) de la loi Normale aux points *x*
  p=cdf('Normal',x,mu,sigma)
- L'inverse de la fonction de répartition (inverse cumulative distribution function) pour la probabilité p
  x=icdf('Normal',p,mu,sigma)

2.13 Box plot



MATLAB

boxplot(x))

2.14 Données vs distributions

Pourquoi choisit-on telle ou telle distribution pour modéliser une variable aléatoire?

- Données de base: $Y_i \sim N(\mu_i, \sigma_i)$ avec $0 \leqslant i \leqslant n$
- Linéaire: $a_j = \sum_i x_{ij} Y_i \sim N\left(\mu = \sum x_{ij}\mu_i,\ \sigma = \sqrt{\sum x_{ij}\sigma_i^2}\right)$
- Moyenne: $\sqrt{n}\ \left(\frac{\bar{Y}-\mu}{s}\right) \sim T(\nu)$
- Fonction quadratique: $(a_j)^2 \sim \chi^2(w_j)$
- Rapport fonctions quadratiques: $\frac{(a_j)^2}{(a_i)^2} \sim F\left(w_j, w_i\right)$

2.15 Distribution $t(\nu)$ de Student et intervalle de confiance

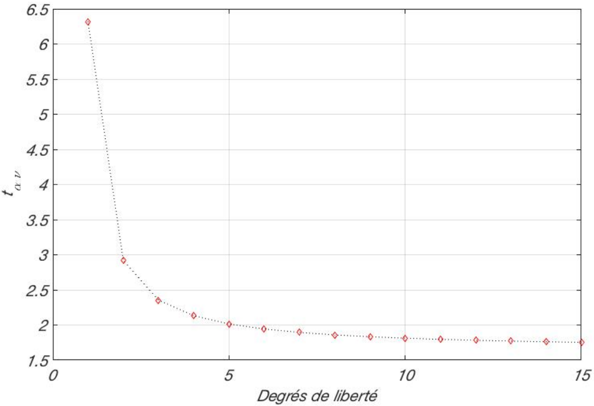
- Distribution très importante pour l'échantillonnage
- Publiée par William Gosset, 1908, (Guinness)
- Par définition, si $\begin{cases} Z \sim N(0,1) \\ U \sim \chi^2(\nu) \end{cases}$ alors

$$T = \frac{Z}{\sqrt{U/\nu}} \sim t(\nu)$$

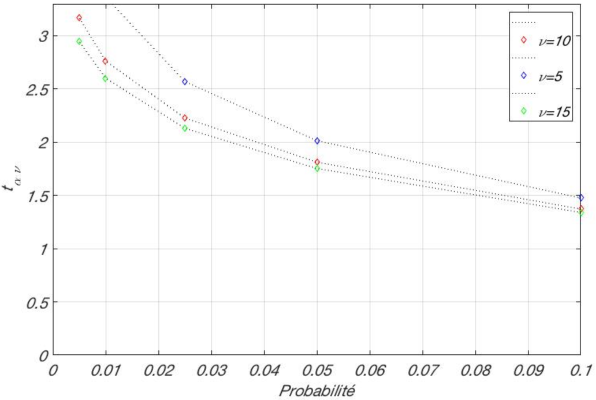
- Donc si $\bar{x}$ est la moyenne des mesures, μ l'espérance (inconnue) de la population de référence, s l'écart type et ν le nombre de degrés de liberté, alors :

$$\frac{\bar{x} - \mu}{s/\sqrt{\nu}} \sim t(\nu) \qquad (2.5)$$

2.16 Influence du degré de liberté sur $t_{0.05}^{\nu}$

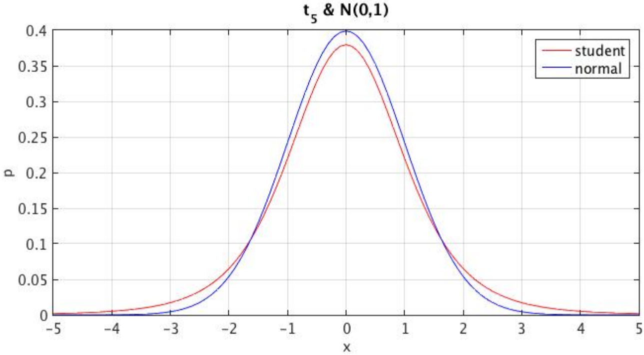


2.17 Influence de la probabilité sur t_{α}^{ν}



2.18 Distribution Normale vs Student

$$T(\nu) \approx N(0,1), \nu \leqslant 30$$



2.18 La distribution t_ν de Student

MATLAB

- Génération de nombre aléatoire selon la loi t_ν pour ν degrés de liberté:
  X=trnd(nu,Ni,Nj)
- Densité de probabilité (probability density function) de la loi *t* aux points *x* pour ν degrés de liberté:
  p=tpdf(x,nu)
- Fonction de répartition (cumulative density function) de la loi t_ν pour ν degrés de liberté aux points *x*:
  p=tcdf(x,nu)   *p=tcdf(x,nu,'upper')*
- Inverse de la fonction de répartition (inverse cumulative distribution function) pour la probabilité *p* et ν degrés de liberté:
  x=tinv(p,nu)

2.19 La covariance

- $cov(X, Y)$ est une mesure de la dépendance entre deux variables X et Y
- Elle s'écrit aussi S_{xy}
- L'unité de la covariance est le produit des unités des deux variables
- Elle se calcule de la manière suivante

$$S_{xy} = \frac{1}{N} \sum_{i=1}^{N} (x_i - \bar{x})(y_i - \bar{y}) \tag{2.6}$$

2.20 Le principe de la variance

De la définition 2.3 on comprend que la variance est une somme de carrés, donc pour une combinaison linéaire, avec *a* et $b \in \mathbb{R}$, on obtient les résultats suivants:

$$
\begin{aligned}
var(a\,Z + b) &= a^2\ var(Z) && (2.7)\\
var(Z_1 + Z_2) &= var(Z_1) + var(Z_2) + 2\ cov(Z_1, Z_2) && (2.8)
\end{aligned}
$$

Si Z_1 et Z_2 sont des variables aléatoires indépendantes, $cov(Z_1, Z_2) = 0$, donc:

$$var(a\,Z_1 + b\,Z_2) = a^2\ var(Z_1) + b^2\ var(Z_2) \tag{2.9}$$

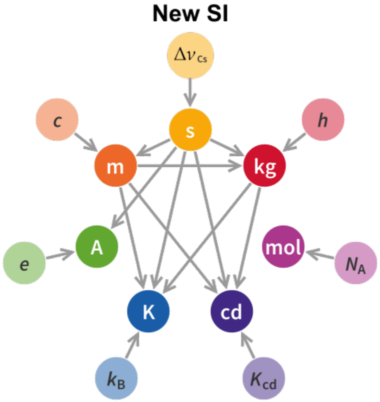
Cas particulier important:

$$var(Z_1 - Z_2) = var(Z_1) + var(Z_2) \tag{2.10}$$

3. Systèmes Physiques

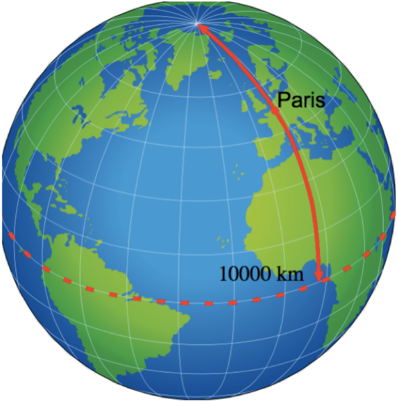
3.1 Système d'unités

- **MKSA** : mètre, kilogramme, seconde, ampère
- **SI**: mètre, kilogramme, seconde, ampère, *Kelvin, mol, candela*

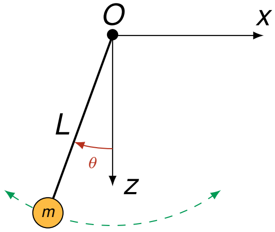


Définition originelle du mètre basée sur la dimension de la Terre

$10'000$ *km* correspond au demi méridien de Paris



Définition originelle de la seconde



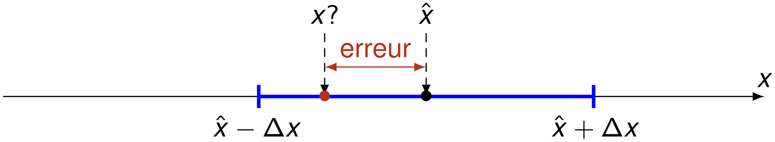
- Pulsation d'un pendule: $\omega = \sqrt{\frac{g}{L}}$
- Période: $T = \frac{2\pi}{\omega}$
- Période d'un pendule de $2m$: $T = \frac{2\pi}{\sqrt{g}} = 2.006$
- 1 seconde correspond à une demi période d'un pendule de 1 métre

Définition de la seconde

Un courant d'un ampère correspond au transport d'une charge électrique d'un coulomb par seconde à travers une surface (section de fil, électrolyte, tube à vide, etc.).

3.2 Erreurs et incertitudes

- Le mot "erreur" n'a pas le même sens dans le langage courant, qu'en science expérimentale où il signifie "incertitude"
- Aucune mesure n'est absolument exacte et donc la valeur vraie est (en général) inconnue
- L'objectif est de tendre à des erreurs aussi petites que possible et d'avoir une estimation de leur amplitude

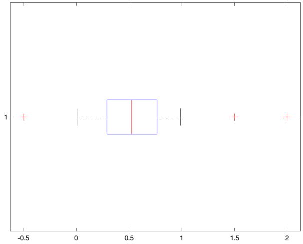


- Un résultat expérimental $\hat{x}$ est incomplet sans une estimation d'erreur:
  1. un intervalle $\pm \Delta x$
  2. la probabilité de cet intervalle, $p(\hat{x} - \Delta x \geqslant x \geqslant \hat{x} + \Delta x)$

3.3 Types d'erreur: erreur accidentelle

- Erreur de protocole
- Erreur de lecture
- Panne d'un élément

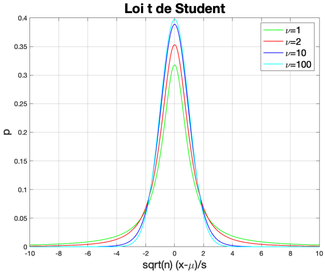
⇒ Analyse visuelle, détection des *outlyer*, élimination des données erronées et répétitions de la mesure



3.4 Types d'erreur: erreur aléatoire

- décrite par une distribution,
- fréquemment la distribution normale

$\Rightarrow$ répliquer la mesure (loi de Student: $N \approx 10$)



3.5 Types d'erreur: erreur systématique

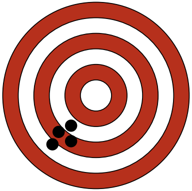
- Qualité de la technologie et de l'étalonnage
- Qualité du protocole et de l'opérateur
- Phénomènes physiques perturbant la mesure

$\Rightarrow$ Améliorer l'étalonnage, contrôler l'expérience

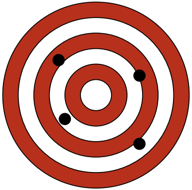
$\Rightarrow$ Les bons instruments de mesure sont calibrés par rapport aux normes maintenues par des bureaux nationaux, voire internationaux, de poids et mesure.

3.6 Erreurs aléatoires vs systématiques

Précision vs exactitude



faibles erreurs aéatoires et fortes erreurs systématiques



fortes erreurs aéatoires et faibles erreurs systématiques

⇒ Vérification et étalonnage systématique des instruments

3.7 Précision d'un voltmètre numérique

Exemple des indications du fabriquant

MODEL	FUNCTION	RANGE	ACCURACY	RESOLUTION
M-3860D M-3850D M-3870D	DC VOLTAGE	400 mV 4 V 40 V 400 V	± 0.3% of rdg +1 dgt	100 µV 1 mV 10 mV 100 mV
		1000 V	± 0.5% of rdg +1 dgt	1 V
M-3850D M-3870D	AC VOLTAGE	400 mV 4 V 40 V 400 V	± 0.8% of rdg +3 dgt	100 µV 1 mV 10 mV 100 mV
		750 V	± 1.0% of rdg +3 dgt	1 V
3860D	AC VOLTAGE (True rms)	400 mV 4 V 40 V	± 0.8% + 3 dgt (± 2.5% + 5 dgt)	100 µV 1 mV 10 mV
		400 V 750V	± 1.0% of rdg +3 dgt	100 mV 1 V

Note: Impedence of AC Voltage True rms (M-3860D)

1. 40Hz to 20KHz for 400mV, 4V, 40V & 200V
2. 40Hz to 1KHz for above 200V to 750V

3.8 La détermination de l'incertitude

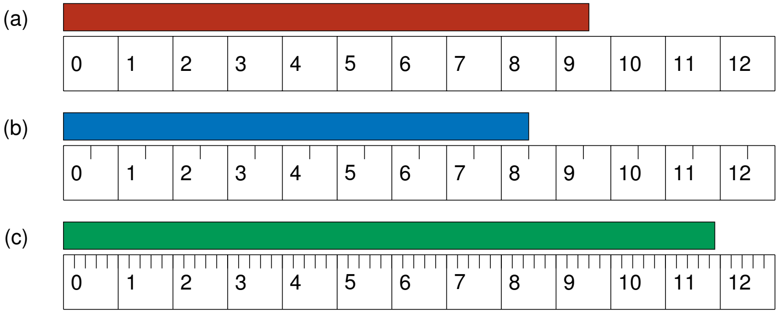
La détermination de l'incertitude est composées de trois éléments

1. La précision de l'instrument
2. L'incertitude estimée
3. L'incertitude statistique lorsqu'il y a des mesures répétées

3.9 La précision de l’instrument

- Pour la précision de l’instrument, il faut considérer:
  - La **plus petite division** (The least count) de l’instrument. Un *mètre* a des divisions espacées d’au moins 1,0 mm.
  - La **précision de l’instrument**, (Instrument Limit of Error, ILE), est la précision à laquelle un appareil de mesure peut être lu. La précision est toujours égale ou inférieure à la plus petite division.
- La précision de l’instrument est généralement considérée comme égale à la plus petite division ou une fraction (1/2, 1/5, 1/10) de celle-ci.
- Aucune règle stricte: utiliser le bon sens.
- Affichages digitaux: la précision est $1/2$ du dernier digit
- Pour certains appareils, la précision est donnée sous forme d’une **tolérance** ou d’un pourcentage.

3.10 Lecture sur une échelle



	Division	Précision	Lecture
(a)			
(b)			
(c)			

3.11 L'incertitude estimée

L'incertitude estimée est généralement plus grande que la précision de l'instrument. Elle dépend de la sensibilité de l'instrument:

- *Exemple 1*: un balance avec des divisions de 0.1 g, mais qui ne réagit pas à moins de 1 g

$\Rightarrow$ L'incertitude sera estimée à ± 0.5 g

- *Exemple 2*: mesure de la longueur focale d'une lentille avec une règle dont la précision est de 0.5 mm. Cependant la position de l'écran peut de 1 cm sans changer la netteté de l'image

$\Rightarrow$ L'incertitude sera estimée à ± 0.5 cm

3.12 Chiffres significatifs et arrondis

Principe:

1. Arrondir l'incertitude à un ou deux chiffres significatifs
2. Arrondir le résultat en gardant le même nombre de décimales
3. Indiquer l'unité

Illustration:

- Mesures effectuées avec un voltmètre
- Sensibilité: 0.02 V
- Moyenne calculée: 12.14286 V
- Écart-type: 0.07313 V

a) ☐ 12.14286 V

b) ☐ (12.14 ± 0.02) V

c) ☐ 12.14286 $V \pm 0.07313$

d) ☐ 12.143 ± 0.073 V

e) ☐ (12.14 ± 0.07)

f) ☐ (12.14 ± 0.07) V

3.13 Intervalle de confiance

Un intervalle de confiance encadre une valeur réelle que l'on cherche à estimer à l'aide de mesures prises par un procédé aléatoire (wikipédia).
A partir de l'équation 2.5, on peut démontrer le théorème de l'intervalle de confiance:

$$P\left(x \in \left[\bar{x} - t_{1-\alpha/2}^{n-1}\sqrt{\frac{S^2}{n}}\ ,\ \bar{x} + t_{1-\alpha/2}^{n-1}\sqrt{\frac{S^2}{n}}\right]\right) = \alpha \qquad (3.1)$$

Avec

- x une variable aléatoire représentant la quantité qu'on cherche à mesurer
- $\bar{x}$ la moyenne de l'échantillon
- n le nombre de mesures
- S^2 la variance sans biais de l'échantillon
- α la probabilité
- $t_{1-\alpha/2}^{n-1}$ le centile de la distribution de Student pour $n-1$ degrés de liberté et la probabilité résiduelle à droite $1-\alpha/2$

3.14 Combinaison des erreurs - méthode 1

- On considère une valeur calculée y à partir de plusieurs mesures, $x_1, x_2, \ldots, x_n$
$$y = y(x_1, x_2, \ldots, x_n)$$
.
- L'erreur Δ_y peut être estimée à partir des erreurs des mesures Δ_{x_i}
$$\Delta_y = \sum_{i=1}^{N} \left| \frac{\partial y}{\partial x_i} \Delta_{x_i} \right| \tag{3.2}$$
- Le résultat expérimental pourra alors être représenté par $y = \hat{y} \pm \Delta_y$
- Il est important que chaque erreur estimée corresponde à un niveau de confiance similaire aux autres et qu'il soit reporté.

3.14 Combinaison des erreurs - méthode 1

Erreurs relatives et absolues, calculées avec la méthode 1, pour les quatre opérations arithmétiques

fonction	erreur absolue	erreur relative
$y = x_1 + x_2$	$\Delta y = \lvert\Delta x_1\rvert + \lvert\Delta x_2\rvert$	
$y = x_1 - x_2$	$\Delta y = \lvert\Delta x_1\rvert + \lvert\Delta x_2\rvert$	
$y = x_1 \cdot x_2$	$\Delta y = \lvert x_2 \, \Delta x_1\rvert + \lvert x_1 \, \Delta x_2\rvert$	$\frac{\Delta y}{y} = \left\lvert\frac{\Delta x_1}{x_1}\right\rvert + \left\lvert\frac{\Delta x_2}{x_2}\right\rvert$
$y = \frac{x_1}{x_2}$	$\Delta y = \left\lvert\frac{\Delta x_1}{x_2}\right\rvert + \left\lvert\frac{-x_1}{x_2^2}\Delta x_2\right\rvert$	$\frac{\Delta y}{y} = \left\lvert\frac{\Delta x_1}{x_1}\right\rvert + \left\lvert\frac{\Delta x_2}{x_2}\right\rvert$

3.15 Combinaison des erreurs - méthode 2

- Soit une valeur calculée y à partir de plusieurs mesures, $x_1, x_2, \ldots, x_n$:

$$y = y(x_1, x_2, \ldots, x_n)$$

.

- L'écart-type s_y peut être calculé à partir des écarts-type s_{x_i}:

$$s_y = \sqrt{\sum_{i=1}^{N} \left(\frac{\partial y}{\partial x_i}\right)^2 s_{x_i}^2} \tag{3.3}$$

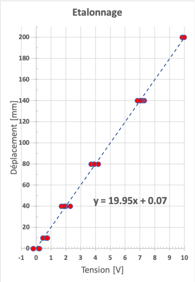
- L'intervalle de confiance pour la probabilité α pourra alors être calculé avec l'équation 3.1 et le résultat expérimental pourra alors être représenté par

$$y = \hat{y} \pm t_{1-\alpha/2}^{n-1} \sqrt{\frac{S^2}{n}}$$

3.16 Exemple d'un capteur de déplacement

- Un capteur de déplacement a une courbe d'étalonnage linéaire, obtenue à partir de 30 mesures, $\ell = a_o + a_1 U$ avec $a_o = 0 \pm 1 mm$ et $a_1 = 20 \pm 0.2 mm/V$

- On cherche l'erreur autour de la valeur $U = 5.0 V$ sachant que l'écart-type de la mesure de la tension est $s_U = 0.01 \ V$



- Selon l'équation 3.3, on peut écrire

$$\begin{aligned}
s_\ell^2 &= \left(\frac{\partial \ell}{\partial a_o} s_{a_o}\right)^2 + \left(\frac{\partial \ell}{\partial a_1} s_{a_1}\right)^2 + \left(\frac{\partial \ell}{\partial U} s_U\right)^2 \\
&= (1 \times s_{a_o})^2 + (U \times s_{a_1})^2 + (a_1 \times s_U)^2 \\
&\approx (1)^2 + (5 \times 0.2)^2 + (20 \times 0.01)^2 = 2.04 \ mm^2 \\
s_\ell &\approx 1.4 \ mm
\end{aligned}$$

  - Donc l'intervalle de confiance à 95% est $\pm s_\ell \times t_{0.025,28} \frac{1}{\sqrt{30}} = \pm 1.4 \times 2.05 \times 0.18 \approx \pm 0.52 mm$