

**Pour les enseignants, les étudiants
et les passionnés !**

Calcul d'incertitudes

Version Courte

**Population et
échantillons**

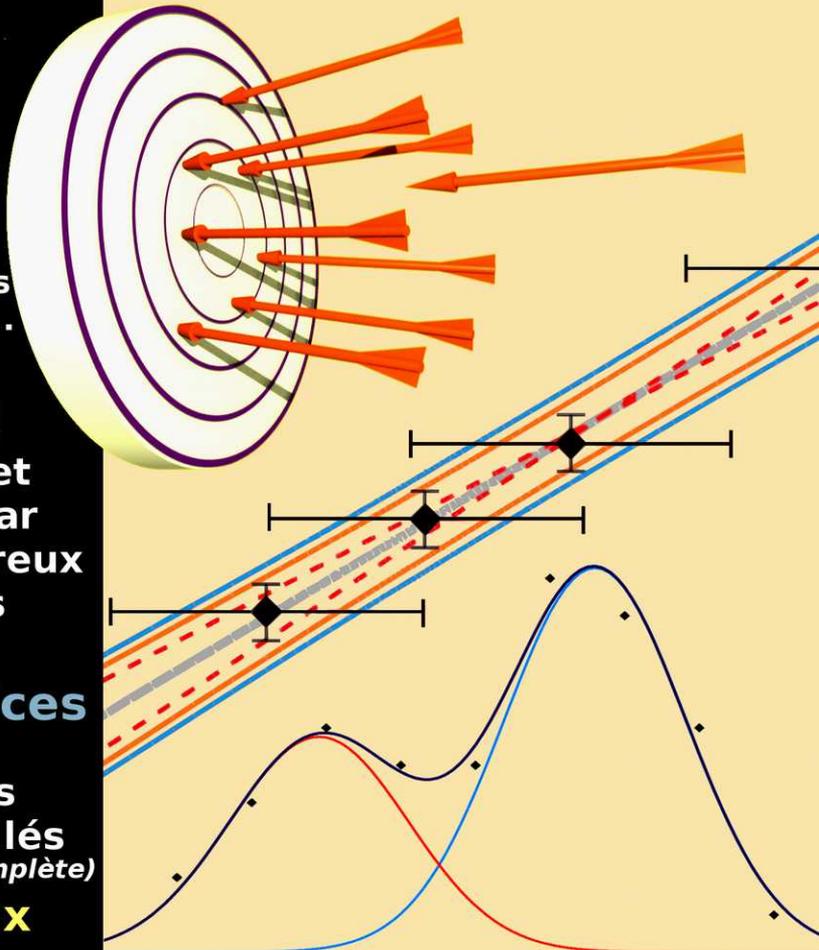
**Sources
des incer-
titudes**

**Régression
linéaire
avec ou
sans barres
d'erreurs ...**

**Cours
complet et
illustré par
de nombreux
exemples**

**+Exercices
avec les
corrigés
détaillés
(en version complète)**

**+Travaux
Pratiques**



Mathieu ROUAUD

Calcul d'incertitudes

*Application
aux sciences expérimentales*

Mathieu ROUAUD

Professeur Agrégé de Sciences Physiques
en classes préparatoires aux grandes écoles d'ingénieurs
Diplômé en Physique Théorique

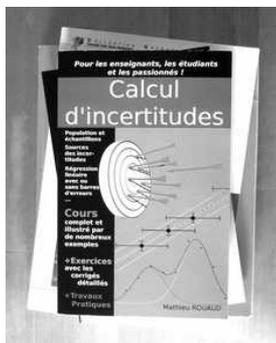
Pour un meilleur partage de la connaissance et l'accès au plus grand nombre, le livre est en licence libre, le livre numérique est gratuit et pour minimiser le coût de la version papier, il est imprimé en noir et blanc et sur papier économique.

**Livres complets
numérique et papier,**

avec tous
les exercices corrigés :
sur www.lulu.com

Pour contacter l'auteur :
ecrire@incertitudes.fr
Boudiguen 29310 Querrien

**Cours particuliers
de mathématiques, physique et
chimie sur demande**



Ce livre est sous licence Creative Commons Attribution-Non Commercial 3.0.



Vous êtes libres :



de reproduire, distribuer et communiquer
cette création au public ,



de modifier cette
création .

Selon les conditions suivantes :



Attribution. Vous devez citer le nom de l'auteur original de la manière indiquée par l'auteur de l'œuvre ou le titulaire des droits qui vous confère cette autorisation (mais pas d'une manière qui suggérerait qu'ils vous soutiennent ou approuvent votre utilisation de l'œuvre).



Pas d'Utilisation Commerciale. Vous n'avez pas le droit d'utiliser cette création à des fins commerciales.

- A chaque réutilisation ou distribution de cette création, vous devez faire apparaître clairement au public les conditions contractuelles de sa mise à disposition. La meilleure manière de les indiquer est un lien vers cette page web : <http://creativecommons.org/licenses/by-nc/2.0/deed.fr>
- Chacune de ces conditions peut être levée si vous obtenez l'autorisation du titulaire des droits sur cette œuvre.
- Rien dans ce contrat ne diminue ou ne restreint le droit moral de l'auteur ou des auteurs.

Date de parution : juillet 2013

Révision : janvier 2017

ISBN 978-2-9549309-0-9

Ouvrage basé sur les deux premiers chapitres prolongés et augmentés du livre *Probabilités, statistiques et analyses multicritères*.

Avant-propos

Cet ouvrage se veut accessible et pédagogique. Il est le fruit d'interrogations personnelles sur la nature probabiliste des mesures en sciences. Dans un cursus classique ces aspects ne sont pas, ou peu, abordés. Il est important que les fondements expérimentaux et pratiques des sciences soient complémentaires d'une science au tableau en cours magistraux. Il existe une beauté scientifique qui naît de l'interaction entre la théorie et l'expérience.

Tout en introduisant les principes fondamentaux de la statistique, cet ouvrage explique comment déterminer les incertitudes dans différentes situations expérimentales. Beaucoup d'exemples sont issus de cours et travaux pratiques réalisés en math sup.

Bonne lecture !

Remerciements :

Je remercie Éric NOIZET (professeur Agrégé de Chimie en prépa) et Grégoire BONNET (ingénieur charpentier) pour leurs multiples apports à la clarté pédagogique de l'ouvrage. Un grand merci à Reine pour sa relecture précise et consciencieuse. Pleins de mercis, à Aurélien SEMACH (étudiant) et, aux enseignants de sciences-physiques Françoise MARCADET (pour ses contributions en métrologie) et Julien BONVALET.

Merci à la vie et à tous ceux qui m'ont précédé.

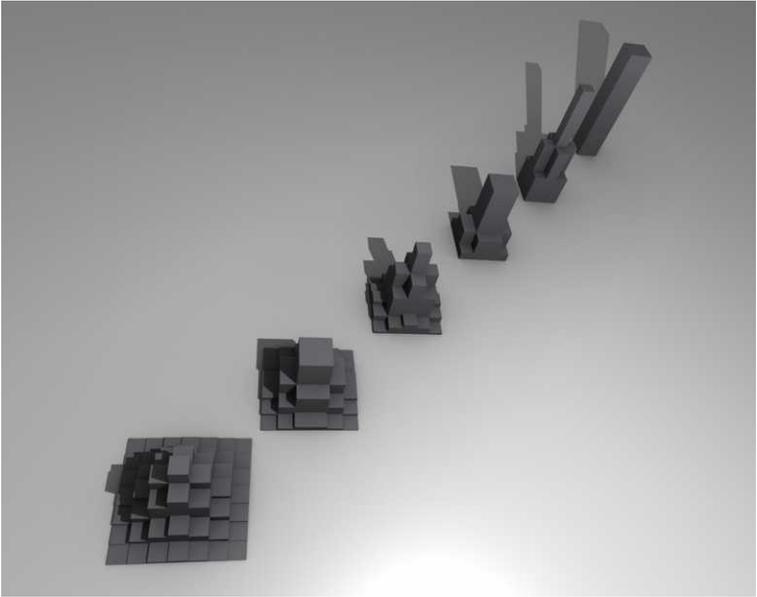


Table des matières

I. VARIABLE ALÉATOIRE.....	1
A. Grandeurs et mesures.....	1
B. Centre d'une distribution.....	1
C. Dispersion d'une distribution.....	3
D. Exemples de distributions.....	4
E. Théorème central limite.....	7
1) Population et échantillons.....	7
2) Le théorème central limite.....	10
3) Coefficient de Student et incertitude.....	12
4) Exemples.....	15
F. Distribution de Gauss.....	19
1) Définition d'une distribution continue.....	19
2) Courbe de Gauss.....	20
3) Loi normale standard.....	23
G. Test d'hypothèse.....	24
H. Test du Khi-deux.....	30
I. Sources des incertitudes.....	33
J. Exercices.....	37
II. CORRÉLATIONS ET INDÉPENDANCES.....	48
A. Coefficient de corrélation.....	48
B. Formule de propagation des incertitudes.....	53
1) Formule de propagation des écart-types.....	53
2) Calcul d'incertitude.....	54
C. Régression linéaire.....	59
1) Principe et formules.....	59
2) Détermination du zéro absolu.....	63
3) Régression avec barres d'erreurs.....	65
4) Linéarisation.....	68

5) Comparaison des méthodes.....	69
D. Régression généralisée.....	76
1) Principe.....	76
2) Régression polynomiale.....	78
3) Régression non linéaire.....	81
E. Exercices.....	87
III. LOIS DE PROBABILITÉS.....	103
A. Lois discrètes.....	104
1) Loi binomiale.....	104
2) Loi géométrique.....	105
3) Loi de Poisson.....	107
B. Lois continues.....	109
1) Loi uniforme.....	109
2) Loi exponentielle.....	111
3) Loi normale.....	112
4) Loi de Student.....	112
5) Loi du Khi-Deux.....	113
C. Fonctions de variables à densité.....	115
D. Simulation numérique.....	118
E. Exercices.....	121
IV. ESTIMATEURS.....	127
A. Qualité d'un estimateur.....	127
1) Biais.....	127
2) Risque.....	128
B. Construction d'estimateurs.....	130
1) Méthode des moments.....	130
2) Méthode du maximum de vraisemblance.....	134
C. Estimation par intervalle.....	138
D. Exercices.....	147

V. COMPLÉMENTS.....	153
A. Mesure avec une règle.....	153
B. Métrologie.....	165
C. Thermodynamique.....	173
D. Indépendance des variables.....	179
VI. DEVOIRS.....	181
A. Devoir Suricate.....	181
B. Devoir Narval.....	184
VII. TRAVAUX PRATIQUES.....	187
A. Mesure d'un indice lumineux.....	187
B. Le miroir sphérique.....	191
C. Relation de conjugaison d'une lentille.....	194
D. Dioptries et lentilles minces sphériques.....	196
VIII. OUTILS MATHÉMATIQUES.....	198
IX. CORRECTIONS.....	203
X. Bibliographie / Sources / Logiciels / Illustrations.....	204
XI. TABLES / Index.....	209
A. Loi normale centrée réduite.....	209
B. Coefficients de Student.....	210
C. Valeurs critiques de Khi-deux.....	211



I. VARIABLE ALÉATOIRE

A. Grandeurs et mesures

Soit X une variable aléatoire et n réalisations $\{x_i\}$ de cette variable.

Nous pouvons simplement estimer une grandeur classique : par exemple, combien y-a-t-il de jours dans une semaine ? La réponse est sans ambiguïté. Par contre pour une grandeur statistique l'approche est plus subtile. Imaginons des étudiants qui font des expériences de calorimétrie pour mesurer la capacité thermique de l'eau¹. Les différents groupes mesurent les valeurs suivantes : $\{5100; 4230; 3750; 4560; 3980\}$ J/K/kg. Que vaut alors la capacité ? Nous donnerons dans ce chapitre une réponse à cette question. Elle sera de nature probabiliste.

B. Centre d'une distribution

Nous cherchons une caractéristique du centre de la distribution des observations $\{x_i\}$. Il en existe plusieurs, le *mode*, par exemple, est facile à déterminer, il s'agit de la valeur la plus représentée (illustrations page 4). Nous avons

1 PHYSIQUE : Quantité d'énergie à fournir à un kilogramme d'eau pour que sa température s'élève de 1°C. L'eau emmagasine ainsi de l'énergie et peut la restituer par la suite en diminuant sa température. Tables : $c_{\text{eau}} = 4180$ Joules par degré Celsius et par kilogrammes.

aussi la *médiane* qui correspond à la valeur qui sépare la distribution en deux parties égales. Mais la plus utilisée est la *moyenne* qui représente au mieux le centre d'une distribution :

$$\bar{x} = \frac{x_1 + x_2 + \dots + x_i + \dots + x_n}{n} \quad \text{soit} \quad \boxed{\bar{x} = \frac{\sum_{i=1}^n x_i}{n}} \quad 2$$

Pour la capacité thermique de l'eau nous obtenons :

$$\bar{c} = \frac{5100 + 4230 + 3750 + 4560 + 3980}{5} = 4324 \text{ J/K/kg}$$

Nous avons considéré la moyenne arithmétique. Nous aurions pu prendre la moyenne géométrique :

$$\bar{x} = \sqrt[n]{\prod x_i}$$

Par exemple, pour deux vitesses de 20 m/s et 40 m/s, la moyenne géométrique est $\sqrt{20 \text{ m/s} \cdot 40 \text{ m/s}} \approx 28,3 \text{ m/s}$ alors que la moyenne arithmétique est de 30 m/s. Dans la pratique on constate que la moyenne arithmétique est mieux adaptée.

2 MATH : se dit «la moyenne de x est égale à la somme de l à n des x_i , le tout divisé par n ». Pour la moyenne géométrique nous considérons la racine $n^{\text{ième}}$ du produit des x_i . \bar{x} , "x moyen", se dit aussi "x barre".

C. Dispersion d'une distribution

Il s'agit d'estimer ce que nous pourrions aussi appeler la largeur d'une distribution. La grandeur la plus simple à déterminer est l'*étendue*, différence entre les valeurs maximale et minimale. Mais celle-ci est très sensible aux valeurs extrêmes qui ne sont pas toujours représentatives, et peuvent même parfois être absurdes.

Dans les faits, la grandeur la plus utilisée est l'*écart-type* :

$$s = \sqrt{\frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n-1}}$$

Pour l'écart-type de la capacité thermique de l'eau nous obtenons :

$$s_c = \sqrt{\frac{(5100-4324)^2 + (4230-4324)^2 + (3750-4324)^2 + (4560-4324)^2 + (3980-4324)^2}{4}}$$

soit $s_c \simeq 530 \text{ J/K/kg}$

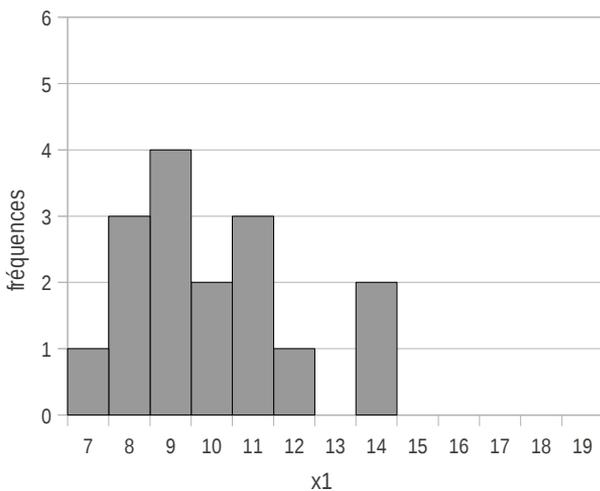
Nous pourrions aussi considérer l'*écart moyen* par rapport à la moyenne (voir l'exercice 1).

Pour l'écart-type si nous divisons par n au lieu de $n-1$, nous obtiendrions l'*écart quadratique moyen*. Le choix de l'écart-type sera justifié par la simplicité des formules qui en découleront (justification complète p129). De plus nous travaillons souvent avec n grand et la différence entre les deux types d'écart quadratiques est alors minime.

D. Exemples de distributions

Cas 1 :

	X_i
X_1^1	11
X_1^2	9
X_1^3	10
X_1^4	14
X_1^5	11
X_1^6	8
X_1^7	9
X_1^8	12
X_1^9	7
X_1^{10}	8
X_1^{11}	8
X_1^{12}	9
X_1^{13}	11
X_1^{14}	14
X_1^{15}	10
X_1^{16}	9



moyenne = 10

mode= 9

médiane= 9,5

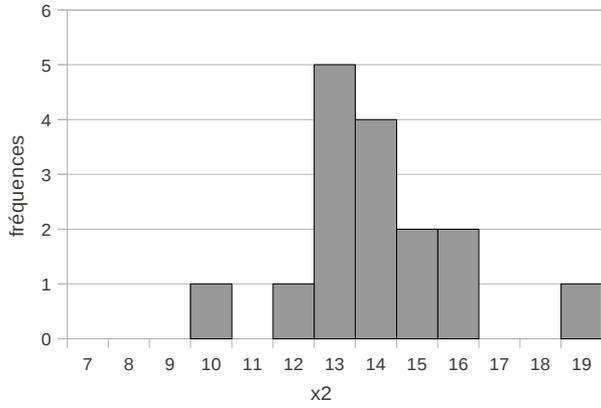
écart-type= 2,07

étendue= 7

écart quadratique moyen= 2,00

Cas 2 :

	X_2
X_2^1	15
X_2^2	13
X_2^3	12
X_2^4	13
X_2^5	14
X_2^6	13
X_2^7	16
X_2^8	19
X_2^9	13
X_2^{10}	14
X_2^{11}	10
X_2^{12}	16
X_2^{13}	14
X_2^{14}	15
X_2^{15}	13
X_2^{16}	14



moyenne = 14

mode= 13

médiane= 14

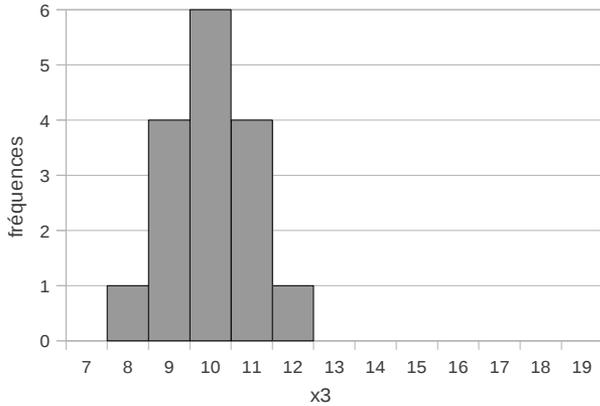
écart-type= 2,00

étendue= 9

écart quadratique moyen= 1,94

Cas 3 :

	X_3
X_3^1	10
X_3^2	10
X_3^3	12
X_3^4	11
X_3^5	9
X_3^6	8
X_3^7	10
X_3^8	9
X_3^9	9
X_3^{10}	11
X_3^{11}	9
X_3^{12}	11
X_3^{13}	10
X_3^{14}	10
X_3^{15}	11
X_3^{16}	10



moyenne = 10

mode= 10

médiane= 10

écart-type= 1,03

étendue= 4

écart quadratique moyen= 1,00

La moyenne n'est pas toujours la valeur la plus représentée (*cas 1 et 2*) et elle peut même dans certains cas être absente. Dans le *cas 3* la courbe est symétrique ce qui implique l'égalité de la médiane et de la moyenne.

Sur les trois exemples certaines valeurs sont représentées plusieurs fois, on parle alors de la *fréquence* f_i d'une valeur

x_i . Nous avons $n = \sum_{i=1}^c f_i$, où c correspond au nombre de

valeurs de x_i différentes auxquelles nous attribuons une fréquence (dans la suite, c sera aussi le nombre de classes).

La moyenne et l'écart-type peuvent alors s'exprimer ainsi :

$$\bar{x} = \frac{\sum_{i=1}^c f_i \cdot x_i}{n} = \sum_{i=1}^c \frac{f_i}{n} \cdot x_i \quad s = \sqrt{\frac{\sum_{i=1}^c f_i (x_i - \bar{x})^2}{n-1}}$$

Parfois on rassemble les valeurs par *classe*, par exemple si nous nous intéressons à la taille des habitants d'une ville nous pouvons rassembler tous les habitants qui ont une taille comprise entre 160 cm et 170 cm dans le même ensemble appelé classe. Leur nombre dans cette classe est la fréquence (ou effectif de la classe) et la valeur est prise égale au milieu de la classe, ici 165 cm (démarche illustrée dans l'exercice 5).

Plus la courbe est ramassée sur son centre plus l'écart-type est faible (sur le *cas 3* l'écart est deux fois plus faible que sur les *cas 1 et 2*).

E. Théorème central limite

1) Population et échantillons

Considérons une ville d'un million d'habitants. Pour sonder la population nous pouvons interroger un échantillon de seulement mille personnes tirées au hasard. A partir de cet échantillon de $n=1000$ individus, nous pouvons, grâce à la statistique, avoir des informations sur la population toute entière. Plus la taille de l'échantillon est grand, plus les résultats seront précis. Nous appelons \bar{x} la moyenne de l'échantillon et s son écart-type. Pour la population nous notons μ (lettre grec mu) la moyenne et σ (sigma) l'écart-type. Plus l'échantillon est grand, plus les valeurs de \bar{x} et de s de cet échantillon sont amenées à se rapprocher de celles μ et σ de la population.

Comme dans le cas des sondages d'opinion avec des échantillons de l'ordre de mille personnes, si nous mesurons la taille de mille habitants qui sont choisis au hasard parmi la population d'une ville d'un million d'habitants, la moyenne de la taille sur cet échantillon a de fortes chances d'être proche de celle sur l'ensemble de la population, mais n'a aucune raison de lui être égale.

Illustrons maintenant par le jeu du lancer de pièces. A chaque lancer nous obtenons une réalisation de la variable aléatoire *pile* ou *face*. Ici la population est infinie, nous pouvons répéter le lancer indéfiniment et avoir une infinité de mesures. De plus, les probabilités étant connues, nous pouvons déterminer à l'avance les caractéristiques de la population.

Tout d'abord, quand la taille de l'échantillon devient très grande et donc infinie, l'échantillon s'identifie à la population : $\mu = \lim_{n \rightarrow \infty} \bar{x}$ ³.

Ensuite nous voyons apparaître la notion de probabilité :

$p_i = \lim_{n \rightarrow \infty} \frac{f_i}{n}$ où p_i est la probabilité de réalisation de l'événement x_i .

D'après la formule page 6 nous avons ainsi l'expression de la moyenne pour la population : $\mu = \sum p_i \cdot x_i$.

Nous avons $\sum p_i = 1$, car nous considérons tous les événements possibles (1=100%).

Nous associons $x_0=0$ à l'observation *face*, et $x_1=1$ pour *pile*. Comme la pièce est équilibrée $p_0=p_1=1/2=0,5=50\%$ et $\mu=p_0 \cdot x_0 + p_1 \cdot x_1$. Nous avons jugé inexistant l'événement *la pièce reste sur la tranche*.

De même nous avons : $\sigma = \lim_{n \rightarrow \infty} s$ et en prenant la limite de la formule pour s page 6 nous obtenons $\sigma = \sqrt{\sum p_i \cdot (x_i - \mu)^2}$ (avec pour n grand, $n-1$ pris égal à n).

Au final : $\mu=0,5$ et $\sigma=0,5$.

3 MATH : se lit « μ est égal à la limite de \bar{x} quand n tend vers l'infini».

Prélevons un échantillon en lançant neuf pièces :

{0; 1; 1; 0; 1; 1; 1; 0; 0}.

Nous avons alors $\bar{x} \simeq 0,56$ et $s \simeq 0,53$.

Supposons que nous tirions de nombreuses fois neuf pièces aléatoirement au sein de la même population. A chaque fois nous aurions un résultat différent pour \bar{x} .

Par exemple pour deux autres échantillons prélevés :

{1; 1; 0; 1; 1; 0; 1; 1; 0} avec $\bar{x} \simeq 0,67$ et $s \simeq 0,50$

et {0; 1; 0; 0; 0; 0; 1; 0; 0} avec $\bar{x} \simeq 0,22$ et $s \simeq 0,44$

Quelle serait alors la répartition de l'ensemble de ces résultats ? (appelée la distribution d'échantillonnage)

Les valeurs obtenues pour les échantillons sont en général différentes de celles de la population, mais plus l'échantillon est grand plus il est probable que les valeurs soient proches de celle de la population. Cas d'un échantillon de taille $n=50$ où $\bar{x} \simeq 0,520$ et $s \simeq 0,505$:

{00100011111110010111110001001101101000110000011101}

2) Le théorème central limite

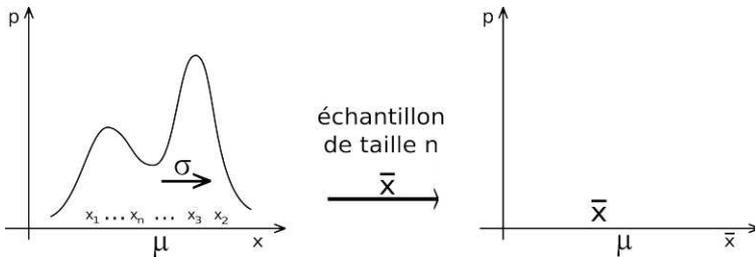
THÉORÈME CENTRAL LIMITE :

Nous prélevons au sein d'une population des échantillons aléatoires de taille n , la moyenne de l'échantillon \bar{x} varie autour de la moyenne de la population μ avec un écart-type égal à σ/\sqrt{n} , où σ est l'écart-type de la population.

Quand n croît la distribution d'échantillonnage de \bar{x} est de plus en plus concentrée autour de μ et devient de plus en plus proche d'une distribution de Gauss.

Nous décrirons prochainement ce qu'est une distribution de Gauss, encore appelée loi normale, dans l'immédiat nous considérerons simplement une courbe en cloche. C'est un théorème très important, **quelque soit** la forme de la distribution de la population, la distribution d'échantillonnage tend vers une **gaussienne**, et sa dispersion est donnée par le théorème central limite.

Illustrons sur des schémas :



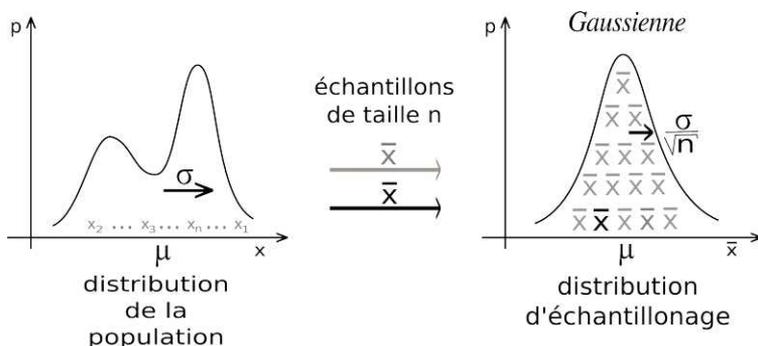
A gauche nous avons la probabilité p d'un événement x (distribution de la population). Par exemple, pour une po-

pulation d'un million d'habitants, pourrait être ici représenté la probabilité p qu'ils aient une taille donnée x . Si nous pouvions mesurer la taille de tous les habitants, nous pourrions déterminer exactement leur taille moyenne μ et son écart-type σ . C'est d'un point de vue pratique difficile, ou même impossible, alors nous décidons de mesurer la taille de mille habitants seulement. Pour que ce soit caractéristique nous tirons ces mille personnes au hasard.

Nous obtenons mille mesures de tailles de x_1 à x_{1000} . De cet échantillon de taille $n=1000$ nous calculons une moyenne \bar{x} et un écart-type s . Nous pensons que \bar{x} est proche de μ , mais à la fois il n'y a aucune raison qu'il soit égale à μ . Nous plaçons cette valeur de \bar{x} sur la partie droite de la figure page 10.

Nous prenons un nouvel échantillon aléatoire de mille personnes et nous plaçons un nouveau \bar{x} .

Nous répétons ensuite cette opération un grand nombre de fois. Nous voyons à droite la distribution des échantillons obtenue :



3) Coefficient de Student et incertitude

Le théorème central limite s'applique dans la limite des grands nombres. Dans le cas particulier où la distribution de la population est normale nous pouvons l'appliquer dès n petit grâce aux coefficients de Student t .

Intervalle de fluctuation :

Si μ et σ sont connus la distribution d'échantillonnage est aussi gaussienne et les fluctuations statistiques attendues se situent avec une probabilité de $p\%$ entre $\mu - t_{\infty} \sigma / \sqrt{n}$ et $\mu + t_{\infty} \sigma / \sqrt{n}$.

Les valeurs de t sont lues dans la table page 210.

Intervalle de confiance :

Dans le cas du calcul d'incertitudes μ et σ ne sont pas connus et nous les estimons à partir de l'échantillon avec \bar{x} et s . Du fait d'une statistique faible il y a alors un élargissement connu donné par la loi de Student :

$$\mu = \bar{x} \pm t \cdot \frac{s}{\sqrt{n}}$$

Le coefficient de Student t dépend de n et inclut la *confiance* que nous voulons donner au résultat. Si la confiance est de 95%, nous avons 95 chances sur 100 que μ soit compris entre $\bar{x} - t \cdot s / \sqrt{n}$ et $\bar{x} + t \cdot s / \sqrt{n}$.

Nous reconnaissons ici la notion d'*incertitude* Δx ⁴:

$$x = \bar{x} \pm \Delta x \quad \text{avec} \quad \Delta x = t \cdot \frac{s}{\sqrt{n}}$$

Δx est aussi appelée l'*incertitude absolue* et $\Delta x / |\bar{x}|$ l'*incertitude relative*.

Reprenons l'expérience de calorimétrie décrite page 1, supposons que nous voulions maintenant connaître la capacité thermique de l'eau avec une confiance de 95%. Comme souvent en sciences expérimentales nous considérons la distribution des données gaussienne, en effet de par l'influence de nombreux facteurs indépendants sur la valeur des grandeurs mesurées nous nous attendons, toujours en vertu du théorème central limite, à avoir normalité des fluctuations.

Nous trouvons dans la table pour quatre *degrés de liberté* ($ddl=n-1$) $t=2,78$.

D'où : $c = \bar{c} \pm t \cdot s_c / \sqrt{n} = 4320 \pm 660 \text{ J/K/kg}$ à 95%.

Ici suite à la dispersion des valeurs mesurées par les étudiants $\Delta c / \bar{c} \approx 15\%$. Les mesures en calorimétrie ne sont pas très précises. La valeur attendue, ici connue, est bien dans l'intervalle : $3660 < 4180 < 4980$.

En sciences expérimentales nous nous efforçons de quantifier l'ensemble des phénomènes naturels. Mais, par la nature même de la démarche expérimentale, les différents paramètres qui permettent de décrire une situation ne sont

4 MATH : se lit "delta x".

pas parfaitement connus. Nous n'avons pas simplement une valeur numérique associée à chaque caractéristique, mais un intervalle pour une confiance donnée. En toute rigueur, à toute grandeur expérimentale doit être associée son incertitude avec sa confiance.

Cas particuliers :

Cas des grands nombres : la taille de l'échantillon n est suffisamment grande pour pouvoir appliquer directement le théorème de la limite centrée. La distribution d'échantillonnage est normale indépendamment de la distribution de la population. Nous n'avons pas à nous soucier de la loi de Student qui de toute façon s'identifie à une loi de Gauss dans ce cas.

Cas des petits nombres et normalité des données : nous appliquons la loi de Student comme précédemment.

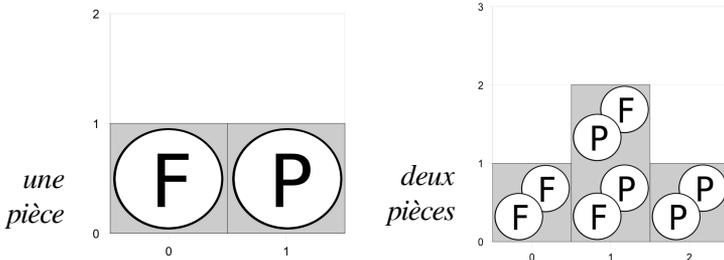
Cas des petits nombres et non normalité des données : par exemple, nous constatons en calculant les coefficients d'asymétrie et d'aplatissement des données qu'elles ne correspondent pas une distribution normale. Nous sommes bloqué et nous devons faire une étude au cas par cas. Par exemple pour des données qui obéissent à une loi uniforme nous avons l'intervalle de fluctuation donné p12 qui fonctionne dès $n=2$ (se montre en s'appuyant sur les données de l'article *Mesure avec une règle* p153) . Par contre pour un loi de Bernouilli de paramètre $p=0,24$ et un échantillon de taille $n=2$, l'intervalle de fluctuation à 50% comporte 0% des valeurs... Un cas plus complexe p145 montre pour $n=12$, par comparaison avec une simulation numérique, que le théorème central sous-estime l'intervalle de confiance.

4) Exemples

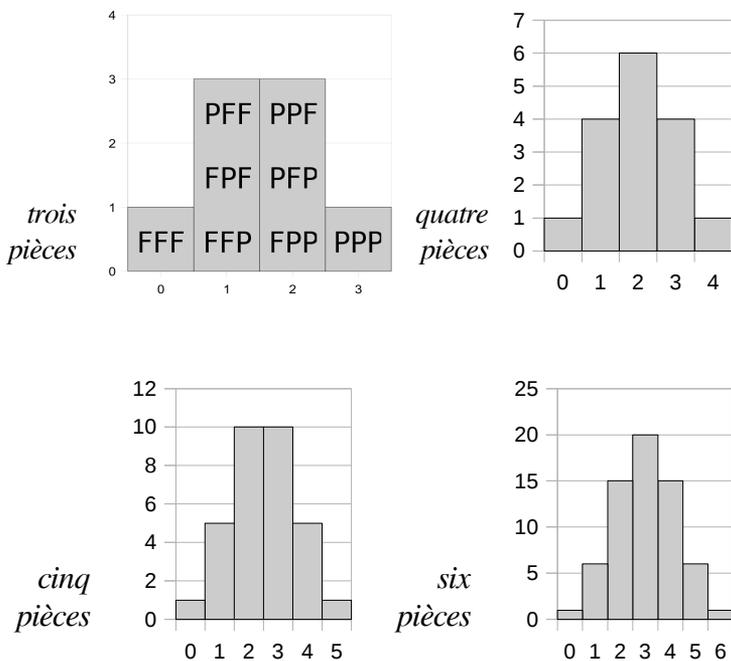
Concrètement un grand nombre de facteurs aléatoires vont influencer sur la mesure d'une grandeur, facteurs indépendants, qui, quelque soient leurs natures vont au final générer une distribution gaussienne des valeurs. Prenons deux exemples, le lancer de pièces et celui de dés à six faces.

Pour les pièces nous comptons le nombre de piles à chaque lancer de plusieurs pièces. Nous dénombrons le nombre de possibilités pour un nombre de piles donné. Pour une pièce, une possibilité pour zéro pile (face, F) et une possibilité pour un pile (P).

Pour deux pièces, une possibilité pour zéro pile (F F), deux possibilités pour un pile (F P ou P F) et une possibilité pour deux piles (P P). Plus le nombre de pièces lancées simultanément est grand, plus nous tendons vers une distribution gaussienne.



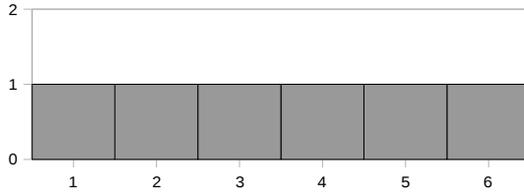
Nous pouvons obtenir en ordonnée la probabilité, en divisant par le nombre de possibilités 2^n , et en abscisse la moyenne pour chaque pièce, en divisant par le nombre de pièces n . Pour $n=1$ nous avons alors la distribution de la population et ensuite les distributions d'échantillonnage pour différentes valeurs de n .



De même pour les dés nous énumérons les possibilités pour leur somme et nous tendons aussi vers une gaussienne.

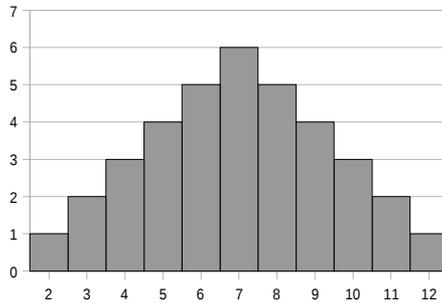
Pour un seul dé, la somme correspond tout simplement à la valeur du dé. Nous avons une possibilité pour chaque valeur :

un dé :

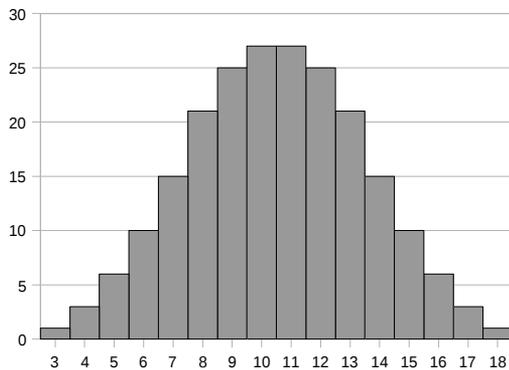


Pour deux dés, il y a une seule possibilité pour que la somme fasse deux : 1 pour le premier dé et 1 pour le deuxième dé. Pour que la somme fasse trois, il y a deux possibilités : 1 puis 2, ou, 2 puis 1. Le plus probable avec deux dés est d'obtenir 7 : (1,6) (6,1) (2,5) (5,2) (3,4) (4,3).

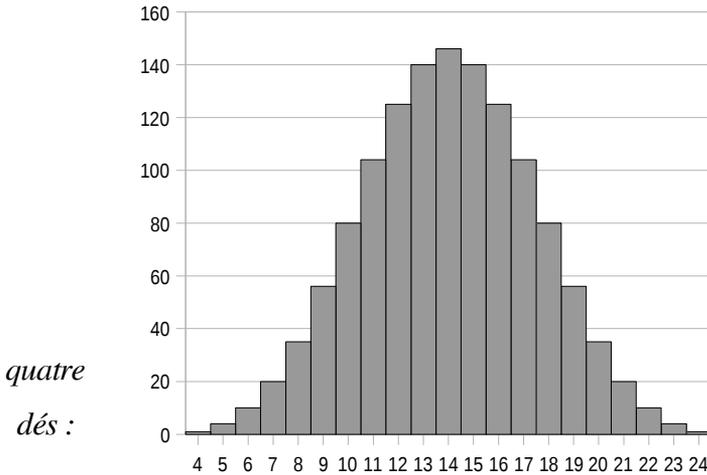
deux dés :



trois dés :



Pour quatre dés nous reconnaissons déjà bien la courbe en cloche et le profil est clairement de type gaussien :



Sur ce dernier exemple vérifions la validité du théorème central limite.

La moyenne pour la population est :

$$\mu_x = (1 + 2 + 3 + 4 + 5 + 6) / 6 = 3,5$$

La moyenne de la distribution d'échantillonnage est bien la même : $\mu_{\bar{x}} = 14 / 6 = 3,5$.

L'écart-type de la population :

$$\sigma_x = \sqrt{\sum p_i \cdot (x_i - \mu)^2} \text{ d'où}$$

$$\sigma_x = \sqrt{1/6 \cdot \{(1-3,5)^2 + (2-3,5)^2 + (3-3,5)^2 + (4-3,5)^2 + (5-3,5)^2 + (6-3,5)^2\}}$$

et $\sigma_x \simeq 1,71$

d'où pour quatre dés $\sigma_{\bar{x}} = \sigma_x / \sqrt{n} = \sigma_x / 2 \approx 0,85$. Or, sur la courbe du dessus, à 40% du maximum (explication page 21), nous avons un écart d'environ 3,5 (entre 3 et 4), soit en moyenne $3,5/4 \approx 0,88$. Il y a bien correspondance.

F. Distribution de Gauss

1) Définition d'une distribution continue

Il existe des grandeurs fondamentalement continues. Le temps en est une. Le temps s'écoule continûment tel un fluide, il ne saute pas d'une valeur à l'autre. S'il semble parfois égrainer les secondes, comme les perles d'un chapelet, cela n'est dû qu'à l'appareil de mesure insuffisamment précis. Par contre quand nous lançons un dé six faces, impossible de dire "j'ai fait 2,35 !", c'est une valeur interdite, seules les valeurs entières de un à six sont permises.

Ainsi, certaines distributions de probabilité sont discrètes et d'autres continues. Pour le lancer de dés la probabilité est discrète : $p_1 = \dots = p_6 = 1/6$ et $\sum_{i=1}^{n=6} p_i = 1$. Le résultat

du lancer n'a que six valeurs possibles.

Par contre si nous nous intéressons à la taille des habitants d'une ville, celles-ci sont toutes possibles, la distribution est continue. Nous parlons alors de densité de probabilité $p(x)$ avec $p(x)dx$ la probabilité d'un évènement d'être entre x et $x+dx$. dx est une petite variation, et $x+dx$ est infiniment voisin de x .

La probabilité que l'évènement se réalise sur l'ensemble des valeurs possibles est de 100% :

$$\int_{x=-\infty}^{+\infty} p(x) dx = 1 \quad 5$$

Calcul de la moyenne et de l'écart-type d'une distribution continue :

$$\mu = \int_{-\infty}^{+\infty} x \cdot p(x) dx \quad \sigma^2 = \int_{-\infty}^{+\infty} (x - \mu)^2 p(x) dx \quad 6$$

2) Courbe de Gauss

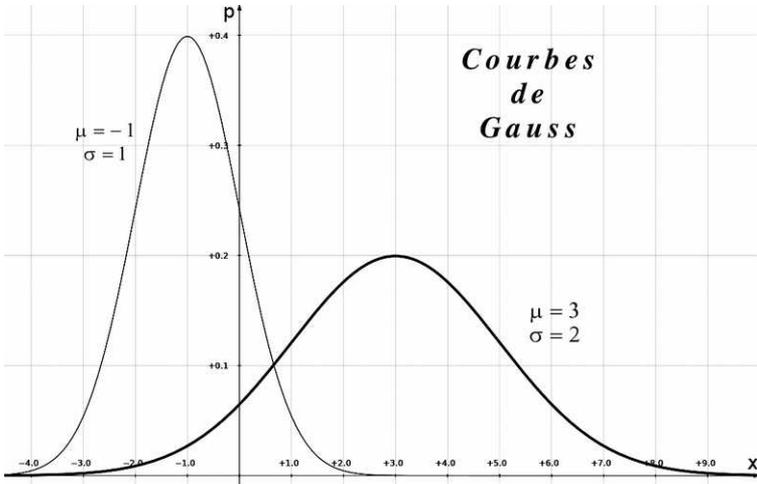
Le théorème central limite s'intéresse au cas où la taille de l'échantillon n est grande, et dans le cas limite où n tend vers l'infini nous considérons une distribution continue. Celle-ci est une gaussienne, l'expression mathématique est connue :

$$p(x) = \frac{1}{\sqrt{2\pi} \cdot \sigma} e^{-\frac{1}{2} \left(\frac{x - \mu}{\sigma} \right)^2}$$

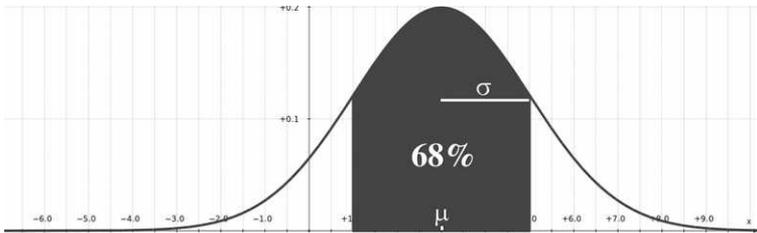
Dans le complément mathématique page 200 différents termes sont justifiés.

-
- 5 MATH : se lit «l'intégrale de p de x , x allant de moins l'infini à $+\infty$, est égale à 1 ».
- 6 MATH : La moyenne est aussi appelée $E(X)$, espérance de X . $\sigma^2 = V(X)$ est appelée variance de la variable aléatoire X . Propriétés : $E(aX+b) = aE(X)+b$, $V(aX) = a^2V(X)$ et $V(X) = E(X^2) - E(X)^2$.

Nous avons représenté, sur le graphe suivant, deux cas :



L'aire totale sous la courbe vaut toujours 1. Mais, si maintenant nous ne nous écartons que d'un écart-type par rapport à la moyenne, nous n'avons que 68% des possibilités :

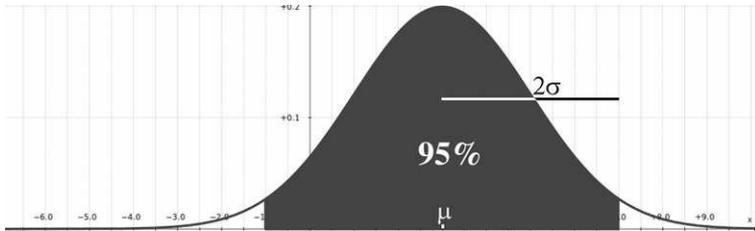


$$\int_{\mu - \sigma}^{\mu + \sigma} p(x) dx = 0,683... \approx 68\%$$

L'écart-type peut s'évaluer ici, en effet à $60\% \cdot p_{max}$:

$$p(\mu \pm \sigma) / p_{max} = 1/\sqrt{e} \approx 0,607$$

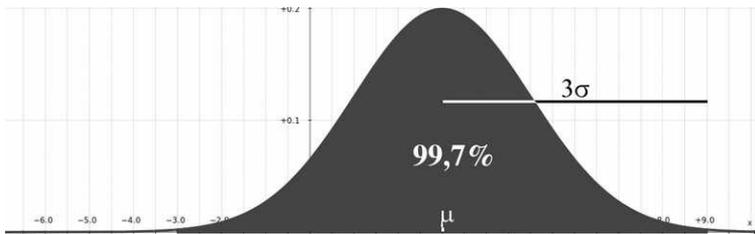
Élargissons à deux écart-types :



$$\int_{\mu - 2\sigma}^{\mu + 2\sigma} p(x) dx = 0,954 \dots \approx 95 \%$$

On dit 95% de confiance, ou, une confiance à deux sigmas.
On travaille souvent avec cette confiance.

Puis pour trois sigmas :



$$\int_{\mu - 3\sigma}^{\mu + 3\sigma} p(x) dx = 0,997 \dots > 99 \%$$

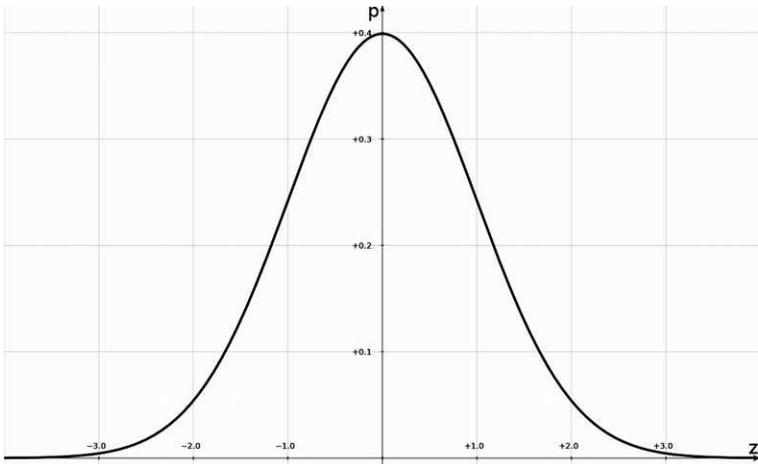
3) Loi normale standard

C'est la distribution normale de Gauss centrée et réduite. Pour le recentrage nous soustrayons la moyenne : $x' = x - \mu$. Pour la réduction nous divisons par l'écart-type :

$$z = \frac{x - \mu}{\sigma}$$

d'où :
$$p(z) = \frac{1}{\sqrt{2\pi}} e^{-\frac{z^2}{2}}$$

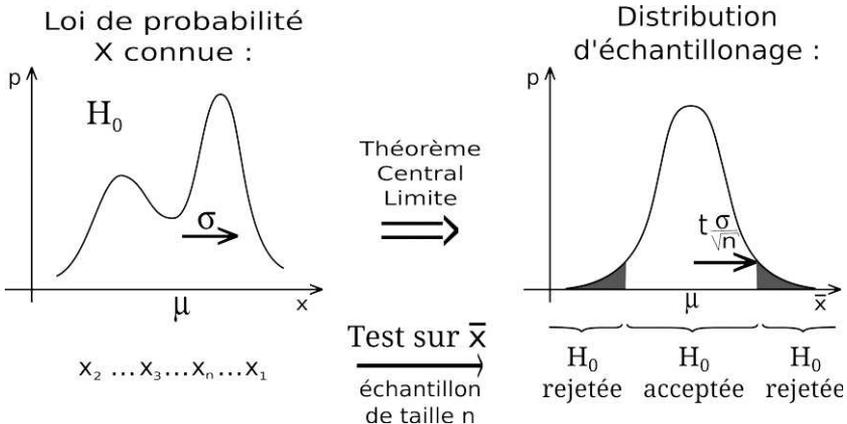
Nous avons alors une distribution normale de moyenne nulle et d'écart-type égale à un :



G. Test d'hypothèse

Nous allons à nouveau utiliser le théorème central limite. Nous l'avons précédemment utilisé pour estimer la moyenne d'une loi de probabilité inconnue. A l'aide d'un échantillon prélevé nous utilisons les propriétés de la distribution d'échantillonnage pour déterminer une grandeur et son incertitude. Ici, pour les tests d'hypothèses nous procédons dans l'autre sens, la loi de probabilité est supposée connue, ainsi la distribution d'échantillonnage est parfaitement déterminée et en prélevant un échantillon nous l'utilisons pour définir un critère de décision permettant d'accepter ou de rejeter l'hypothèse.

En utilisant la moyenne nous testons une hypothèse H_0 faite sur une loi de probabilité X . Soient μ et σ sa moyenne et son écart-type. Nous prélevons au sein de la population un échantillon de taille n suffisamment grand. La valeur moyenne sur l'échantillon est égale à \bar{x} . Si \bar{x} est comprise entre $\mu - t_{\infty} \cdot \sigma / \sqrt{n}$ et $\mu + t_{\infty} \cdot \sigma / \sqrt{n}$ alors l'hypothèse H_0 est acceptée, sinon l'hypothèse est rejetée (test bilatéral).



Nous considérons la valeur t_∞ de la loi normale adaptée à la confiance p désirée.

Nous pouvons utiliser aussi d'autres intervalles caractéristiques de la distribution d'échantillonnage. De manière générale, l'hypothèse H_0 implique une propriété A de la distribution d'échantillonnage. Ici, l'implication n'est pas déterministe mais statistique et la prise de décision procède différemment.

Cas d'un test déterministe : $H_0 \Rightarrow A$

- Si nous observons A alors H_0 ne peut être rejetée.
- Si nous n'observons pas A alors H_0 est rejetée⁷.

Cas d'un test statistique :

Dans $p\%$ des cas : $H_0 \Rightarrow A$

Dans $(1-p)\%$ des cas : $H_0 \not\Rightarrow A$

- Si nous observons A alors H_0 ne peut être rejetée avec $p\%$ de confiance.
- Si nous n'observons pas A alors H_0 est rejetée avec un risque de $(1-p)\%$ de se tromper.

Ce protocole permet de prendre une décision, mais il comporte en même temps des risques de se tromper. Nous pouvons rejeter l'hypothèse alors quelle est vraie ou nous pouvons accepter l'hypothèse alors qu'elle est fausse.

7 Proposition contraposée : si $P \Rightarrow Q$ alors $\bar{Q} \Rightarrow \bar{P}$.

	H_0 est acceptée	H_0 est rejetée
H_0 est vraie	Bonne décision	Mauvaise décision (risque de première espèce α)
H_0 est fausse	Mauvaise décision (risque de deuxième espèce β)	Bonne décision

Nous devons chercher à minimiser les risques liés aux probabilités α et β . Si nous acceptons H_0 , alors α (la probabilité de rejeter l'hypothèse alors quelle est vraie) doit être importante, et β (la probabilité de d'accepter l'hypothèse alors quelle est fausse) doit être faible. Si au contraire nous rejetons H_0 c'est que α est faible et β importante.

Dans le cas où H_0 est vraie nous pouvons faire le calcul du risque, et α vaut tout simplement $1-p$. Par contre si H_0 est fausse, nous ne pouvons faire le calcul de β , à moins de disposer d'une hypothèse alternative H_1 connue.

Pour un test classique on fixe $1-p$ à l'avance, par exemple nous pourrions considérer le test statistiquement discernable au seuil de 5%, et suivant le résultat trouvé nous rejetons ou acceptons l'hypothèse. Nous pouvons aussi calculer les probabilités critiques α et β qui correspondent à la valeur \bar{x} trouvée avec notre échantillon. Nous mesurerons ainsi la crédibilité de H_0 et nous ferons ensuite le choix d'accepter ou pas notre hypothèse.

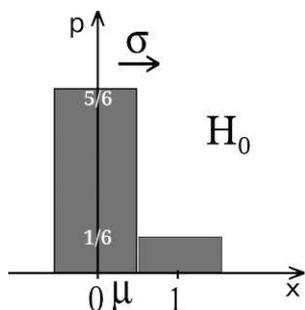
Par exemple imaginons que nous disposions de plusieurs dés. Tous les dés sont équilibrés sauf un, conçu pour

tomber sur le six avec une probabilité double. Malheureusement le dé pipé s'est mélangé aux autres et il n'a pas de signe distinctif. Nous en choisissons un à emmener pour une soirée jeux et nous voulons tout d'abord nous assurer qu'il est bien équilibré.

Nous effectuons 92 lancers :

3151353256243365313441354244652632465436616546
2241154636433165514444241456414316555146362534

Pour l'hypothèse H_0 nous définissons la variable aléatoire discrète X qui vaut zéro lorsqu'on ne fait pas un six et qui vaut un sinon avec $P(X=0)=5/6$ et $P(X=1)=1/6$.



x_i	$x_0 = 0$	$x_1 = 1$
p_i	$p_0 = 5/6$	$p_1 = 1/6$

$$\mu = p_0 x_0 + p_1 x_1$$

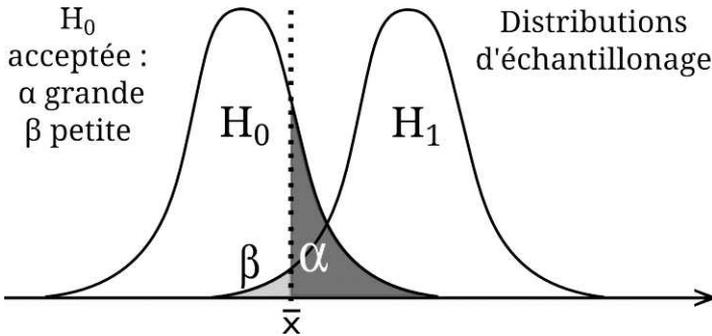
$$\sigma^2 = p_0(x_0 - \mu)^2 + p_1(x_1 - \mu)^2$$

Cette loi a pour moyenne $\mu = 1/6 \approx 0,167$ et pour écart-type $\sigma = \sqrt{5/6} \approx 0,373$.

Dans notre échantillon il y a seize 6 d'où la moyenne $\bar{x} = 16/92 \approx 0,174$ ainsi $\bar{x} - \mu = t_{\infty} \cdot \sigma / \sqrt{n}$ et nous avons un élargissement : $t_{\infty} = (\bar{x} - \mu) \sqrt{n} / \sigma \approx 0,187$.

L'aire sous la courbe de Gauss pour des valeurs supérieures à 0,187 vaut alors 0,43 et $\alpha \approx 43\%$ ⁸.

8 Ici σ est connu et $n=92$ est suffisamment grand utiliser le théorème central limite dans la limite des grands nombres. Pour le calcul sur LibreOffice : 1- LOI.NORMALE.STANDARD(0,187).



Si H_0 est fautive alors H_1 est vraie. Si l'hypothèse alternative H_1 était que le dé n'était pas équilibré, sans spécifier dans quel sens, le test aurait été bilatéral et nous aurions pris pour l'aire α celle des deux queues de distribution à gauche et à droite. Ici le test est unilatéral, si H_0 était fautive la probabilité d'observer six serait doublée et nous observerions des valeurs supérieures.

Pour H_1 nous définissons la variable aléatoire Y avec $P(Y=0)=2/3$ et $P(Y=1)=1/3$. Cette loi a pour moyenne $\mu' = 1/3 \approx 0,333$ et pour écart-type $\sigma' = \sqrt{2}/3 \approx 0,471$. D'où $\bar{x} - \mu' = t_{\infty}' \cdot \sigma' / \sqrt{n}$ et $t_{\infty}' = (\mu' - \bar{x}) \sqrt{n} / \sigma' \approx 3,24$.

L'aire de la queue inférieure vaut alors $\beta \approx 0,06\%$.

Nous pouvons donc très confortablement accepter l'hypothèse que le dé choisi est équilibré, dans le cas de rejet on aurait 43% de chance de se tromper (on essaye en priorité de minimiser cette erreur, classiquement ce n'est qu'en dessous du seuil de 5% que l'on commence à remettre en cause l'hypothèse de départ). Au regard de l'hypothèse alternative il y a moins d'une chance sur 1000 qu'on l'on ai considéré le dé équilibré alors qu'il est pipé (on parle aussi

de puissance du test : $\eta=1-\beta$).

Il est à noter que nous ne calculons jamais la probabilité qu'une hypothèse soit vraie mais la probabilité de la rejeter alors quelle est vraie (risque de première espèce).

Dans le cadre juridique on fait une erreur de première espèce si une personne innocente est condamnée et de deuxième espèce si une personne coupable est relaxée. On demande au jury que la culpabilité soit prouvée au-delà d'un doute raisonnable, si la personne est condamnée α doit être suffisamment petit ^[vi]. Nous cherchons à minimiser la probabilité de le condamner alors qu'il est innocent. Nous ne considérons pas directement la probabilité qu'il soit coupable, la personne jugée est présumée innocente (H_0).

Les exercices 7 et 8 traitent différentes configurations possibles pour ce test.

H. Test du Khi-deux

Il s'agit d'un test d'hypothèse simple basé sur les différences entre les effectifs observés et ceux espérés :

Effectifs observés :

O_1	O_2	...	O_j	...	O_c
-------	-------	-----	-------	-----	-------

Effectifs espérés :

E_1	E_2	...	E_j	...	E_c
-------	-------	-----	-------	-----	-------

Nous calculons la somme des écarts au carré par rapport à la valeur théorique attendue :

$$\chi^2 = \sum_{j=1}^c \frac{(O_j - E_j)^2}{E_j} \quad (\text{nommé "khi-deux"})$$

Nous avons ensuite une table (page 211) qui permet d'estimer la probabilité que l'hypothèse H_0 soit rejetée alors qu'elle est vraie. Suivant la valeur de χ^2 et le nombre de degrés de liberté nous déterminons si l'hypothèse est acceptable. Les degrés de liberté se calculent ainsi :

$$ddl = c - 1 \quad (\text{nombre de valeurs moins une unité})$$

Illustrons avec les expériences réalisées par le botaniste Mendel. Il effectue des croisements entre des plants, ici des hybrides. Il croise entre eux des pois à fleurs roses. Sa théorie lui indique qu'il doit obtenir 25% de pois à fleurs rouges, 25% de pois à fleurs blanches et 50% de pois à fleurs roses. Ceci résulte de la rencontre aléatoire des gamètes. Imaginons qu'il observe sur mille fleurs les valeurs suivantes : 27% de blanches, 24% de rouges et 49% de roses. Doit-il continuer à croire en son hypothèse ?

Effectifs observés :

270	240	490
-----	-----	-----

Effectifs espérés :

250	250	500
-----	-----	-----

d'où

$$\chi^2 = \frac{(270-250)^2}{250} + \frac{(240-250)^2}{250} + \frac{(490-500)^2}{500} \simeq 2,2 \text{ et}$$

$$ddl = 3 - 1 = 2 \text{ .}$$

Donc, d'après la table, il y a plus de 30% de chance que l'hypothèse soit rejetée alors qu'elle est juste.

Nous décidons alors d'accepter l'hypothèse. En général on prend une probabilité critique α de 5%, en dessous de laquelle on envisage de rejeter l'hypothèse.

Le test se généralise facilement pour un tableau :

Effectifs observés :

Effectifs espérés :

$$\left(\begin{array}{cccccc} O_{11} & O_{12} & \dots & O_{1j} & \dots & O_{1c} \\ O_{21} & O_{22} & \dots & \dots & \dots & \dots \\ \dots & \dots & \dots & \dots & \dots & \dots \\ O_{i1} & \dots & \dots & O_{ij} & \dots & \dots \\ \dots & \dots & \dots & \dots & \dots & \dots \\ O_{l1} & \dots & \dots & \dots & \dots & O_{lc} \end{array} \right) \quad \left(\begin{array}{cccccc} E_{11} & E_{12} & \dots & E_{1j} & \dots & E_{1c} \\ E_{21} & E_{22} & \dots & \dots & \dots & \dots \\ \dots & \dots & \dots & \dots & \dots & \dots \\ E_{i1} & \dots & \dots & E_{ij} & \dots & \dots \\ \dots & \dots & \dots & \dots & \dots & \dots \\ E_{l1} & \dots & \dots & \dots & \dots & E_{lc} \end{array} \right)$$

Les ddl tiennent compte du nombre de colonnes c et de lignes l : $ddl = (c-1)(l-1)$

Le χ^2 se calcul avec le même type de formule :

$$\chi^2 = \sum_{(i,j)} \frac{(O_{ij} - E_{ij})^2}{E_{ij}}$$

De plus nous utilisons la même table pour déterminer la validité de l'hypothèse.

I. Sources des incertitudes

Une variable aléatoire a une incertitude d'autant plus faible que la mesure est fidèle, juste et que le système d'acquisition a une bonne résolution.

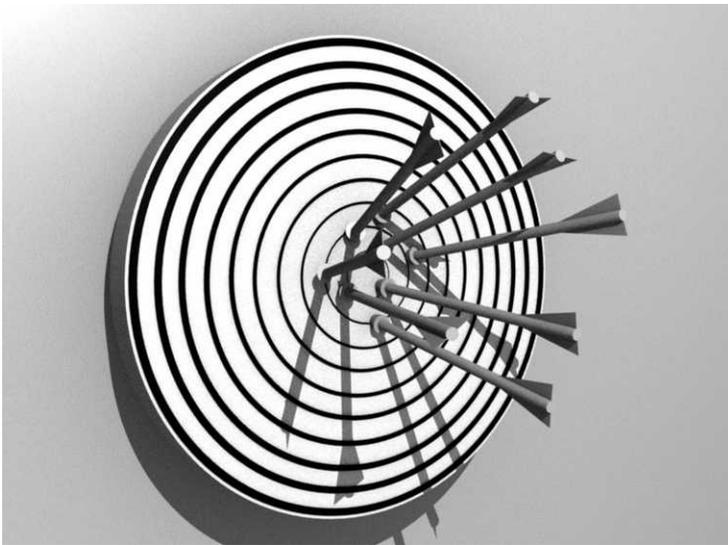
La justesse est assurée par l'absence d'erreurs systématiques. Il peut exister un biais qui rend la mesure inexacte (même si la dispersion est faible). Erreurs de lecture, absence de contrôle et de corrections de facteurs influents, incertitude due à la modélisation, etc. Tous les biais doivent être identifiés et estimés afin d'être ajoutés à la dispersion, le système devient alors juste.

La fidélité provient de la répétabilité et de la reproductibilité des mesures. Les valeurs d'un système fidèle sont peu dispersées. La dispersion peut provenir d'erreurs accidentelles ou d'un phénomène physique par essence aléatoire (comme par exemple la radioactivité). Les expérimentateurs par un travail propre, consciencieux et selon un protocole bien défini et réfléchi, pourront minimiser la dispersion. Les sources peuvent être innombrables, nous essaierons d'en identifier un maximum afin de les évaluer.

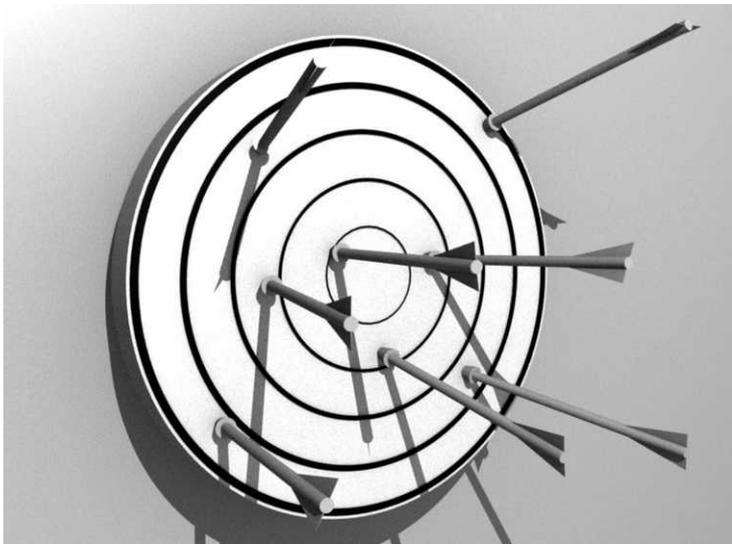
La résolution de l'instrument de mesure dépend de la taille des graduations, du type du vernier ou du nombre de digits de l'affichage. A l'incertitude due à la discrétisation des mesures peuvent s'ajouter d'autres facteurs. Il faudra se référer à la notice ou contacter le fabricant pour connaître au mieux la précision de votre appareil. On peut aussi effectuer un étalonnage avec un instrument de haute précision qui sert de référence.

L'influence de ces différentes sources d'incertitude peut être illustrée par une cible et des flèches. Le centre de la cible correspond à la grandeur à mesurer et les flèches représentent les différentes mesures. Si les flèches, dans leur ensemble, ne sont pas correctement centrées, la justesse n'est pas assurée. Le resserrement des flèches représente la fidélité. La distance entre les cercles sur la cible indique la résolution. La valeur notée est celle du cercle dont la flèche est le plus proche. L'expérimentateur voit les flèches et les cercles, par contre il ne sait pas où est la cible et son centre. Il tient l'arc et son désir d'être au plus proche du centre de la cible montre la qualité et la rigueur de son travail.

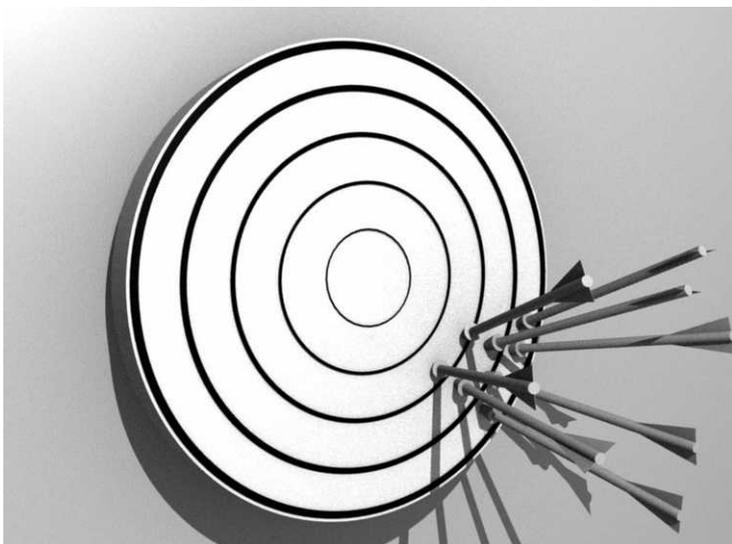
Mesure juste, fidèle et avec une bonne résolution :



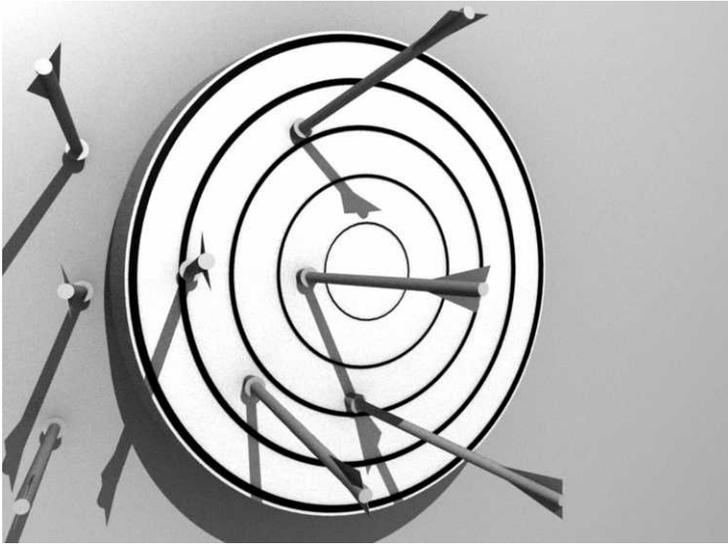
Mesure juste, mais peu fidèle et avec une faible résolution :



Mesure peu dispersée mais avec un biais, et mal résolue :



*Mesure avec un biais, fortement dispersée
et une faible résolution :*



L'écart-type complet sera déterminé à partir des écarts de chaque source en ajoutant les carrés (dus aux compensations explicitées au chapitre 2) :

$$\sigma = \sqrt{\sigma_1^2 + \sigma_2^2 + \sigma_3^2 + \dots}$$

J. Exercices

Exercice 1 : Âges *corrigé en version complète*

Soient les âges des étudiants d'une classe {18; 20; 18; 19; 18; 18; 18; 17; 18; 19; 17; 19; 17; 21; 18}. Déterminez le mode, la médiane, la moyenne arithmétique, la moyenne géométrique, l'étendue, l'écart-type, l'écart quadratique moyen et l'écart moyen⁹.

Exercice 2 : Cartes *corrigé en version complète*

Soit un jeu de 32 cartes. Nous tirons au hasard cinq cartes. Déterminez la probabilité d'avoir un carré d'as, puis celle d'avoir une couleur.

Exercice 3 : Champ de pesanteur *corrigé en version complète*

Des étudiants mesurent l'intensité g du champ de pesanteur terrestre au laboratoire. Les huit binômes mesurent les valeurs suivantes en m/s^2 : 6,20 ; 8,35 ; 13,00 ; 8,37 ; 8,54 ; 9,67 ; 9,75 ; 10,66.

- Quel commentaire général feriez-vous en voyant ces résultats.
- Calculez la moyenne et l'écart-type.
- La dispersion des valeurs donne quelle incertitude sur la moyenne (confiance de 95%) ? Le résultat est-il cohérent avec la valeur attendue ?
- Un neuvième binôme fait une nouvelle mesure dans les mêmes conditions expérimentales. Évaluez la probabilité qu'ils obtiennent un résultat entre 8 et 12 m/s^2 .

9 $\text{Écart moyen} = \left(\sum |x_i - \bar{x}| \right) / n = \left(\sum \sqrt{|x_i - \bar{x}|^2} \right) / n$

Exercice 4 : Ascenseur *corrigé en version complète*

La charge maximale d'un ascenseur est de 300 kg, la masse totale des occupants est de 280 ± 10 kg à σ . Quelle est la probabilité d'être en surcharge ?

Exercice 5 : Devoir *corrigé en version complète*

Pour un devoir les notes obtenues par des étudiants sont les suivantes (notation sur 20) :

10	5	13	7	6	9	5	5	10	15	5	3
15	12	11	1	3	13	11	10	2	7	2	8
2	15	4	11	11	5	8	12	10	18	6	

- Déterminez la moyenne et l'écart-type.
- Réalisez un diagramme avec les notes en abscisse et les fréquences en ordonnée.
- Faire un autre diagramme des fréquences en regroupant les données en classes : $[0, 1, 2]$, $[3, 4, 5]$... , $[18, 19, 20]$. On indiquera aussi les fréquences relatives. Comparez la lisibilité avec le premier diagramme.

Exercice 6 : Yams *corrigé en version complète*

Le Yams se joue avec cinq dés équilibrés à six faces.

- On lance les cinq dès, quelle est la probabilité de faire un Yams directement (mêmes valeurs sur les cinq dés).
- Pour un lancer quelle est la probabilité d'avoir un total inférieur à dix ?

3) Nous effectuons une série de lancers et nous obtenons les totaux suivants : 18, 15, 17, 22, 16, 12, 14, 22, 23, 14, 23, 14, 18, 21, 12, 15, 18, 13, 15, 18, 17, 15, 17, 21, 25, 16, 8, 15, 15, 13.

- a) Calculez la moyenne et l'écart-type.
- b) Quelle est l'estimation de la moyenne avec un intervalle de confiance de 95% ? Est-ce que cela correspond à la valeur théorique ?
- c) Réalisez un graphe des valeurs avec leurs fréquences correspondantes.
- d) Si nous effectuons un nouveau lancer, quelle est, selon vous, la probabilité d'avoir un résultat supérieur à 24 ?

Exercice 7 : Élastiques *corrigé en version complète*

Un fabricant d'élastiques indique, que sur mille élastiques vendus, dix en moyenne ne sont pas fonctionnels. Un acheteur veut tester les lots livrés avant d'accepter la livraison. Il décide de refuser la livraison si le nombre d'élastiques endommagés est trop élevé et il veut avoir moins de 1% de chance de se tromper en réfutant l'indication du fabricant. L'acheteur prélève n élastiques au hasard, à partir de quel nombre n_0 d'élastiques endommagés doit-il refuser la livraison ? Indiquer ce nombre pour trois cas différents : un échantillon de 1000, 200 puis 50 élastiques.

Exercice 8 : Test d'un isolant *corrigé en version complète*

Un fabricant indique pour un panneau isolant en cellulose une conductivité thermique de 0,039 W/m.K. La valeur est certifiée à 5% près. Vous voulez vérifier si c'est vrai, pour cela vous prenez dix panneaux au hasard et vous mesurez leurs

conductivités respectives ($\text{mW}\cdot\text{m}^{-1}\cdot\text{K}^{-1}$) :

39,1	38,8	39,5	39,2	38,9	39,1	39,2	41,1	38,6	39,3
------	------	------	------	------	------	------	------	------	------

Les valeurs sont-elles conformes à celles annoncées par le fabricant (il considère une confiance de 95% sur la marge donnée) ? Pourrait-il, selon vos résultats, annoncer une autre marge ?

Exercice 9 : Pièces *corrigé en version complète*

Nous effectuons un grand nombre de lancers d'une pièce, nous cherchons à déterminer si les probabilités d'obtenir un pile ou un face sont égales. Nous effectuons l'expérience avec trois pièces différentes, sont-elles équilibrées ? (réponses avec une confiance de 95%)

- 1) 42 piles et 58 faces.
- 2) 510 piles et 490 faces.
- 3) 420 piles et 580 faces.

Exercice 10 : Parité *corrigé en version complète*

La parité hommes-femmes est-elle respectée dans les deux chambres et le conseil constitutionnel ?

	Hommes	Femmes
Assemblée nationale	470	107
Sénat	272	76
Conseil constitutionnel	10	2

Exercice 11 : Naissances *corrigé en version complète*

Testons l'hypothèse suivante : les naissances en Suède se répartissent uniformément tout au long de l'année. Supposons que nous avons un échantillon aléatoire de 88 naissances. Les résultats sont regroupés selon des saisons de longueur variable : 27 naissances au printemps (avril-juin), 20 en été (juillet/août), 8 en automne (septembre/octobre) et 33 en hiver (novembre-mars).

Au seuil de 5% l'hypothèse peut-elle être rejetée ?

Sur le même exemple on collecte maintenant un très grand échantillon : 23385, 14978, 14106 et 35804.

Quelle est alors la conclusion ?

Théorie

Exercice 12 : Gaussiennes dans le plan et l'espace *corrigé en version complète*

Gaussienne à une dimension :

Soit la loi de probabilité de Gauss centrée et réduite $p(x)$.

1- Calculez et comparez la moyenne de x et de $|x|$.
Laquelle correspond à la distance moyenne au point d'origine ? Que dire de σ_x et de $\sigma_{|x|}$?

2- Calculez numériquement $P(|x| \leq 1)$, $P(|x| \leq 2)$ et $P(|x| \leq 3)$.

Gaussienne bidimensionnelle :

Soit la loi de probabilité qui généralise une gaussienne à deux dimensions suivante : $p(x, y) = p(x)p(y)$ avec $p(x)$ et $p(y)$ des lois normales centrées et réduites.

Aides :

Soit des intégrales multiples avec des bornes indépendantes des variables et des fonctions continues, les variables sont alors séparables :

$$\iint f(x, y) dx dy = \iint f(x) f(y) dx dy = \int f(x) dx \int f(y) dy$$

Passage en coordonnées polaires :

$$\rho^2 = x^2 + y^2 \text{ et } dx dy = 2\pi \rho d\rho \text{ (symétrie par rotation)}$$

1- Quelle est l'expression de $p(x, y)$? Montrez que

$p(x, y)$ vérifie les deux conditions nécessaires pour une loi de probabilité.

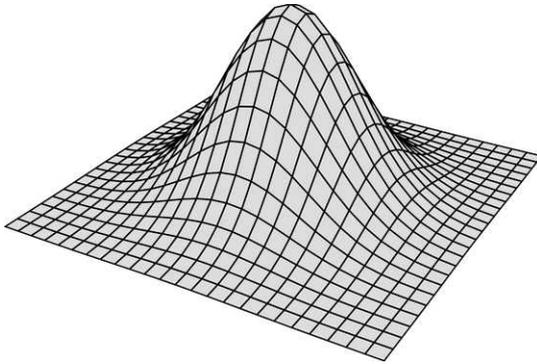
2- En introduisant les coordonnées polaires vérifiez que la probabilité sur tout le plan vaut bien l'unité. Vous exprimerez $p(\rho)$ définie tel que :

$$\iint p(x, y) dx dy = \int p(\rho) d\rho .$$

$p(\rho)$ est la densité de probabilité par rapport à ρ .
 $p(\rho) d\rho$ correspond à la probabilité d'un événement d'être entre ρ et $\rho+d\rho$.

Que valent la moyenne de la distance au point d'origine $\bar{\rho}$ et l'écart-type σ_ρ pour cette loi de probabilité ?

3- Calculez $P(\rho \leq \sigma_\rho)$, $P(\rho \leq 2\sigma_\rho)$ et $P(\rho \leq 3\sigma_\rho)$.



Gaussienne tridimensionnelle :

Soit la loi de probabilité qui généralise une gaussienne à trois dimensions suivante :

$p(x, y, z) = p(x)p(y)p(z)$ avec $p(x)$, $p(y)$ et $p(z)$ des lois normales centrées et réduites.

Aides :

Nous avons avec les intégrales triples les mêmes conditions de séparation des variables qu'avec les intégrales doubles.

Passage en coordonnées sphériques :

$r^2 = x^2 + y^2 + z^2$ et $dx dy dz = 4\pi r^2 dr$
(symétrie sphérique)

1- Quelle est l'expression de $p(x, y, z)$? Exprimez $p(r)$ définie tel que : $\iiint p(x, y, z) dx dy dz = \int p(r) dr$.
 $p(r)$ est la densité de probabilité par rapport à r .
 $p(r) dr$ correspond à la probabilité d'un événement d'être entre r et $r+dr$.

2- Avec cette nouvelle expression vérifiez que la probabilité sur tout l'espace vaut bien l'unité.

3- Que valent la moyenne \bar{r} et l'écart-type σ_r pour cette loi de probabilité ?

4- Calculez $P(R \leq \sigma_r)$, $P(R \leq 2\sigma_r)$ et $P(R \leq 3\sigma_r)$.

5- Comparez les trois cas d'une gaussienne 1D, 2D ou 3D. Vous devez retrouver les résultats du tableau qui suit.

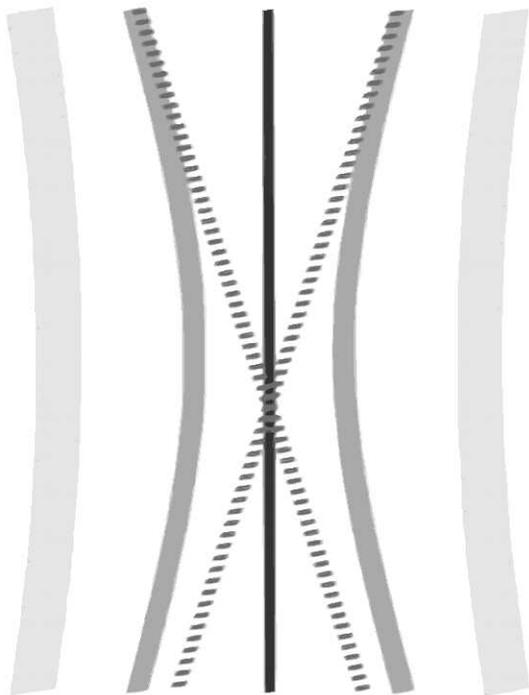
Gaussienne	1D	2D	3D
Distance à l'origine	$ x ^{10}$	ρ	r
Moyenne	$\sqrt{\frac{2}{\pi}}$	$\sqrt{\frac{\pi}{2}}$	$2\sqrt{\frac{2}{\pi}}$
Écart-types	1	$\sqrt{2}$	$\sqrt{3}$
P à σ	68,3%	$1-1/e=63,2\%$	60,8%
P à 2σ	95,4%	98,2%	99,3%
P à 3σ	99,7%	99,988%	99,9994%

Pour vérifier vos calculs sur un tableur vous pouvez utiliser les fonctions suivantes :

sur OpenOffice :

- somme d'une zone sélectionnée * (ex.: B43:B53) : =SOMME(*)
- valeur fixée à la cellule B3 : \$B\$3
- =MOYENNE(*)
- carré : $*^2$
- racine carrée : $*^{(1/2)}$
- =ECARTYPE(*)
- coefficient de Student pour une confiance de 95% et $n=20$: =LOI.STUDENT.INVERSE(0,05;19)
- probabilité que l'hypothèse soit vraie, tableau des valeurs observées * (ex.: B71:E72), celui des valeurs estimées **: =TEST.KHIDEUX(*;**).

¹⁰ MATH : Nous pourrions nous étonner de la différence de l'expression en 1D avec la valeur absolue, ce n'est qu'une question de définitions en coordonnées cylindriques et sphériques. Par exemple, $\rho \in [0; +\infty[$ et $\theta \in [0; 2\pi[$, mais nous pourrions aussi prendre $\rho \in]-\infty; +\infty[$ et $\theta \in [0; \pi[$, la moyenne de ρ serait alors nulle et nous aurions considéré sa valeur absolue.



II. CORRÉLATIONS ET INDÉPENDANCES

Au chapitre précédent nous n'avions qu'une grandeur aléatoire X avec ses n réalisations $\{x_i\}$. Maintenant nous avons plusieurs grandeurs et un nouvel indice permet de les distinguer : X_j et ses mesures $\{x_{jk}\}$. X_j est la $j^{\text{ème}}$ grandeur et x_{jk} est la $k^{\text{ème}}$ observation de cette grandeur. Nous allons nous intéresser aux interactions entre différentes grandeurs.

Pour illustrer, considérons un échantillon de quatre individus qui possèdent trois caractéristiques, la taille X_1 , le poids X_2 et le mois de naissance X_3 . A priori, nous nous attendons à une corrélation entre la taille et le poids : plus on est grand plus on a, en général, une masse importante (corrélation positive). Par contre, nous pouvons penser que le mois de naissance n'a aucune incidence sur le poids et la taille (X_3 non corrélée avec X_1 et X_2).

A. Coefficient de corrélation

Le coefficient de corrélation r permet d'identifier s'il y a une relation linéaire entre deux variables X_i et X_j :

$$r_{ij} = \frac{\sum_k [(x_{ik} - \bar{x}_i)(x_{jk} - \bar{x}_j)]}{\sqrt{\sum_k [(x_{ik} - \bar{x}_i)^2]} \cdot \sqrt{\sum_k [(x_{jk} - \bar{x}_j)^2]}}$$

r varie entre -1 et $+1$. Si $|r|=1$ les variables sont parfaitement corrélées : $r=1$ même sens de variation, $r=-1$ sens opposé. Si $r=0$ il n'y a pas la moindre corrélation, les variables sont parfaitement indépendantes.

Calcul de r_{12} , r_{13} et r_{23} :

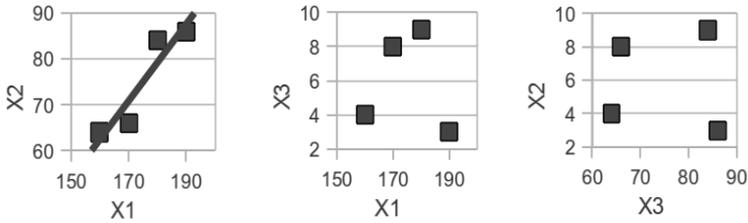
	X_1 (cm)	X_2 (kg)	X_3	$x_1 - \bar{x}_1$	$x_2 - \bar{x}_2$	$x_3 - \bar{x}_3$	$(x_1 - \bar{x}_1)^2$
1	160	64	4	-15	-11	-2	225
2	170	66	8	-5	-9	2	25
3	180	84	9	5	9	3	25
4	190	86	3	15	11	-3	225
\bar{x}	175	75	6			$\Sigma=$	500

et la suite des calculs :

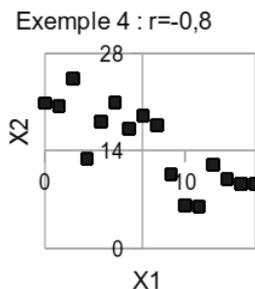
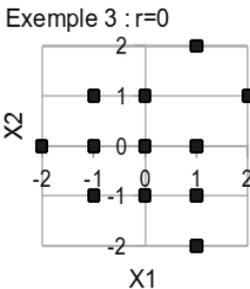
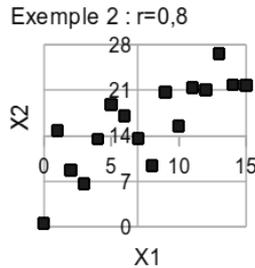
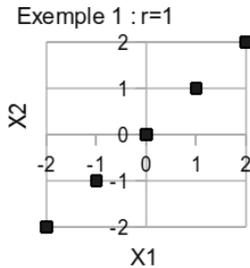
$(x_2 - \bar{x}_2)^2$	$(x_3 - \bar{x}_3)^2$	$(x_1 - \bar{x}_1)$ $\cdot (x_2 - \bar{x}_2)$	$(x_1 - \bar{x}_1)$ $\cdot (x_3 - \bar{x}_3)$	$(x_2 - \bar{x}_2)$ $\cdot (x_3 - \bar{x}_3)$
121	4	165	30	22
81	4	45	-10	-18
81	9	45	15	27
121	9	165	-45	-33
404	26	420	-10	-2

d'où : $r_{12} = \frac{420}{\sqrt{500}\sqrt{404}} \simeq 0,93$, $r_{13} \simeq -0,09$ et $r_{23} \simeq -0,02$.

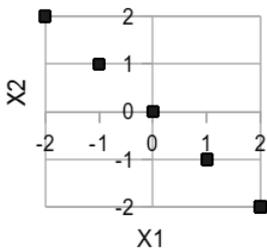
r_{12} est proche de +1, nous avons donc une corrélation positive importante. r_{13} et r_{23} sont proches de zéro : X_3 est indépendante de X_1 et X_2 :



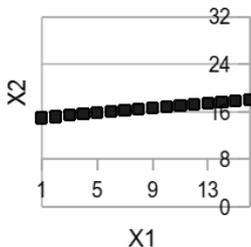
Exemples de nuages de points entre deux variables :



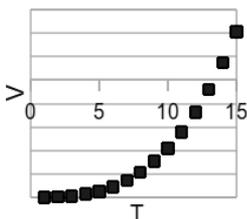
Exemple 5 : $r=-1$



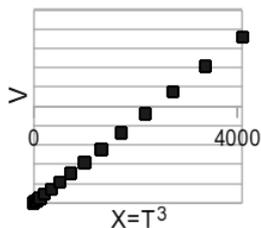
Exemple 6 : $r=1$



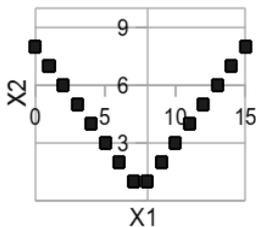
Exemple 7 : $r=0,92$



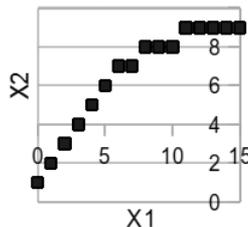
Exemple 8 : $r=1$



Exemple 9 : $r=0$



Exemple 10 : $r=0,94$



Nous voyons sur les exemples 7, 9 et 10 une forte corrélation entre les deux variables. Pourtant le coefficient de corrélation n'est pas aussi proche de -1 ou +1 que nous pourrions l'imaginer, il est même nul sur l'exemple 9. Ceci provient du fait que les corrélations ne sont pas linéaires.

Il peut y avoir des phénomènes de saturation (exemple 10) ou de seuil (un produit peut être bénéfique à faible dose et nocif pour des doses plus importantes - exemple 9). Il est préférable d'avoir identifié ces phénomènes en amont. Pour cela il est important de bien réfléchir à la pertinence des variables que l'on choisit avant de commencer une étude statistique.

Autre exemple : si nous étudions le volume V de différents objets en fonction de leur taille T , nous trouvons une corrélation positive. Mais celle-ci sera bien plus forte entre V et $X=T^3$ (graphes 7 et 8).

Le fait que deux variables soient corrélées n'implique pas nécessairement une relation de cause à effet entre les deux variables (la variation d'une variable entraîne la variation de l'autre). Les variations des deux variables peuvent être attribuables à une cause commune extérieure.

Par exemple, on peut supposer qu'il y a corrélation entre la consommation d'huile à bronzer et celle de crèmes glacées. Il n'y a évidemment aucuns liens de cause à effet entre les deux, mais une cause commune d'ordre météorologique.

Des raisons physiques permettent de montrer une causalité, pas une étude statistique.

B. Formule de propagation des incertitudes

Soit un seau rempli d'un million de grains de sable. La masse d'un grain est de 10 mg à 1mg près. Qu'elle est la masse de sable contenue dans le seau ?

1) Formule de propagation des écart-types

Pour une approche générale du problème, soit une fonction f qui dépend de p variables indépendantes :

$$f(x_1, x_2, \dots, x_j, \dots, x_p)$$

A chacune de ces variables aléatoires est associée une valeur moyenne \bar{x}_j et un écart-type σ_j .

Que valent f et σ_f ?

La statistique donne la réponse et démontre la formule de propagation des écart-types :

$$\sigma_f^2 = \sum_{j=1}^p \left[\left(\frac{\partial f}{\partial x_j} \right)^2 \sigma_j^2 \right] \quad 11$$

11 Nous obtenons la formule de la variance en remplaçant σ^2 par V.

2) Calcul d'incertitude

Pour les incertitudes (définies page 13) nous avons aussi une formule de propagation :

$$\Delta f^2 = \sum_{j=1}^p \left[\left(\frac{\partial f}{\partial x_j} \right)^2 \Delta x_j^2 \right]$$

La formule de propagation des incertitudes n'est pas exacte comme pour les écart-types, mais elle est très pratique et le plus souvent très proche du résultat exact.

Pour notre seau : $M(m_1, m_2, \dots, m_j, \dots, m_p)$

avec
$$M = \sum_{j=1}^p m_j$$

où nous appelons M la masse totale de sable dans le seau, m_j la masse de chaque grain et p le nombre de grains.

$$\Delta M^2 = \sum_{j=1}^p \left(\partial M / \partial m_j \right)^2 \Delta m_j^2$$

$$\begin{aligned} \partial M / \partial m_j &= \partial m_1 / \partial m_j + \dots + \partial m_j / \partial m_j + \dots + \partial m_p / \partial m_j \\ \partial M / \partial m_j &= 0 + \dots + 1 + \dots + 0 = 1 \end{aligned}$$

(les calculs de dérivées partielles sont expliqués page 199)

$$\text{alors} \quad \Delta M^2 = \left(\sum_{j=1}^p 1^2 \right) \Delta m^2$$

d'où $\Delta M^2 = p \cdot \Delta m^2$ avec $\Delta m = \Delta m_j$ quelque soit j .

Finalemnt : $\Delta M = \sqrt{p} \cdot \Delta m = \sqrt{1000000} \cdot 0,001 \text{ g}$.

Le seau pèse donc dix kilos à un gramme près. La précision sur la masse du seau est donc de 0,01%. Naïvement, nous aurions pu penser que l'incertitude globale sur la masse du seau était la somme des incertitudes de chaque grain, nous aurions alors une incertitude absolue d'un kilo et une relative de 10%, ce qui est très différent de la réalité et ignorerait les compensations.

Ici la formule de propagation est très précise car nous avons un très grand nombre de grains. Elle est même exacte, dès les petits nombres, si la distribution de la masse des grains est gaussienne¹².

12 MATH : une combinaison linéaire de grandeurs gaussiennes est elle-même gaussienne (s'applique ici à une somme). Et dans la formule de propagation des incertitudes, si f et les x_i ont des lois de probabilités de même nature, la formule est tout aussi exacte que la formule de propagation des écart-types.

En pratique, certains cas se rencontrent souvent :

- Pour des sommes ou des différences les incertitudes absolues au carré s'ajoutent :

$$\Delta f^2 = \sum_{j=1}^p \Delta x_j^2$$

Par exemple si $d = x_2 - x_1$ avec $\Delta x_2 = \Delta x_1 = 1 \text{ cm}$ alors
 $\Delta d \simeq 1,4 \text{ cm}$.

- pour des produits ou des quotients les incertitudes relatives au carré s'ajoutent :

$$\left(\frac{\Delta f}{f} \right)^2 = \sum_{j=1}^p \left(\frac{\Delta x_j}{x_j} \right)^2$$

Par exemple si $R = U/I$ avec U et I à 1% alors R est connu à 1,4%.

Dans les cas plus complexes, il faut réaliser explicitement le calcul aux dérivées partielles.

A l'aide d'un ordinateur il est aussi possible de faire un calcul numérique. On peut générer des nombres aléatoires ou utiliser des paquets décorrélés globalement gaussiens. Cette dernière méthode est illustrée dans l'exercice 2 de ce

chapitre. Un tableur peut permettre de faire automatiquement les calculs (par exemple feuille 4 du tableur IncertitudesLibres sur www.incertitudes.fr).

On peut aussi donner des méthodes qui donnent des idées générales sur l'incertitude. Elles développent une certaine intuition mais il y a un risque de se tromper.

Par exemple comme les incertitudes s'ajoutent au carré, on peut considérer que l'incertitude la plus grande va l'emporter rapidement sur les autres. Dans l'exemple où $R=U/I$ si U est connu à 1% et I à 0,1% alors R est connu à 1,005% \approx 1%, on peut négliger l'incertitude sur I .

Pour des sommes ou des différences, on considère parfois que le terme qui a le dernier chiffre significatif le moins précis indique la précision du dernier chiffre significatif du résultat. Mais sur notre exemple de la masse du seau rempli de grains de sable, ça ne marche pas : la masse d'un grain est $m = 10$ mg or la masse du seau M est connue au gramme et non au milligramme !

Pour les produits ou des quotients, on considère parfois que le facteur qui a le nombre de chiffres significatifs le plus faible indique le nombre de chiffres significatifs du résultat,

mais là aussi il faut être très attentif.

Une illustration très simple, si $H=2h$ avec $h=5,00m$ (h connu au cm), que vaut H ? D'après la règle du dessus H serait connu avec trois chiffres significatifs, donc $H=10,0m$. H ne serait connu qu'à $10cm$ près, il va de soit que l'on est plus vers le cm ...

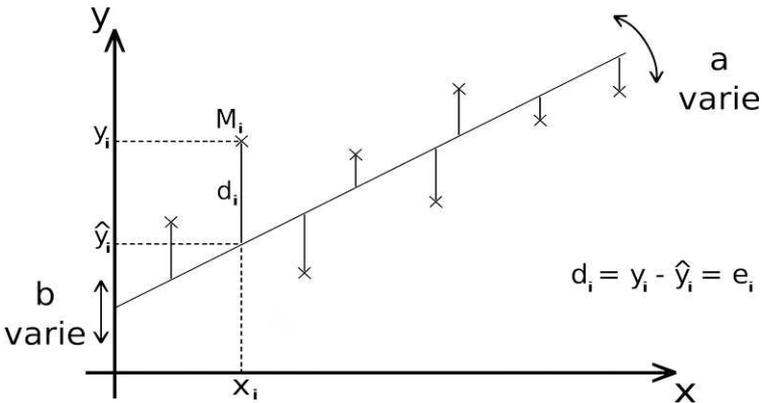
N'utilisez pas ces recettes si vous n'avez pas conscience des pièges possibles.

C. Régression linéaire

Si maintenant nous avons deux variables corrélées nous pourrions vouloir déterminer la relation la plus adaptée entre elles. Les variables aléatoires seront nommées X et Y et nous chercherons la droite qui passe au mieux par le nuage de points $y(x)$. Par exemple, quelle est la relation la mieux adaptée entre la taille X et le poids Y dans notre exemple initial : $y = ax + b$?

Quelles sont les incertitudes Δa et Δb ?

1) Principe et formules



La méthode choisie est celle des moindres carrés : la droite considérée la meilleure est celle qui minimise la somme des carrés des distances à la droite, distances prises selon y (écarts).

L'ensemble des points se note $M_i(x_i, y_i)$. Pour x_i donné,

l'ordonnée estimée sur la droite s'écrit $\hat{y}_i = a x_i + b$.

D'où la somme des distances au carré :

$$\sum d^2 = \sum_i (y_i - \hat{y}_i)^2$$

Les dérivées partielles de cette quantité selon a et b s'annulent pour la meilleure droite et nous obtenons les équations suivantes : $\sum_i (y_i - a x_i - b) x_i = 0$ et

$$\sum_i (y_i - a x_i - b) = 0 .$$

Nous obtenons ainsi les résultats désirés :

$$a = \frac{\bar{x}\bar{y} - \bar{x}\bar{y}}{\bar{x}^2 - \bar{x}^2} \quad \text{et} \quad b = \bar{y} - a\bar{x} .$$

On nomme e_i le résidu tel que $y_i = \hat{y}_i + e_i$.

On trouve les différents écart-types suivants ¹³:

- pour les résidus $s_r = \sqrt{\frac{\sum_i e_i^2}{n-2}}$
- pour la pente $s_a = \frac{s_r}{\sqrt{\sum_i (x_i - \bar{x})^2}}$
- pour l'ordonnée à l'origine $s_b = s_r \sqrt{\frac{\sum x_i^2}{n \sum (x_i - \bar{x})^2}}$

13 Démonstrations p98.

Puis $\Delta a = t_{n-2} s_a$ et $\Delta b = t_{n-2} s_b$.

t_{n-2} : coefficients de Student pour $n-2$ degrés de liberté.

Vous êtes ainsi à même de mener tous les calculs.

Effectuons les calculs pour le poids en fonction de la taille:

$$\bar{x}\bar{y} = (160 \times 64 + 170 \times 66 + 180 \times 84 + 190 \times 86) / 4$$

$$\bar{x}^2 = (160^2 + 170^2 + 180^2 + 190^2) / 4$$

$$a = (13230 - 175 \times 75) / (30750 - 175^2) = 0,84 \quad \text{et}$$

$$b = 75 - 0,84 \times 175 = -72$$

$$s_r = \sqrt{[(64 - (0,84 \times 160 - 72))^2 + (-4,8)^2 + 4,8^2 + (-1,6)^2] / 2} \simeq 5,06$$

$$s_a \simeq 5,06 / \sqrt{(160 - 175)^2 + (-5)^2 + 5^2 + 15^2} \simeq 0,226$$

$$\Delta a \simeq 2,92 \times 0,226 \simeq 0,66 \quad \text{avec une confiance de 90\%}$$

$$s_b \simeq 5,06 \sqrt{(160^2 + 170^2 + 180^2 + 190^2) / [4(15^2 + 5^2 + 25 + 225)]} \simeq 39,7$$

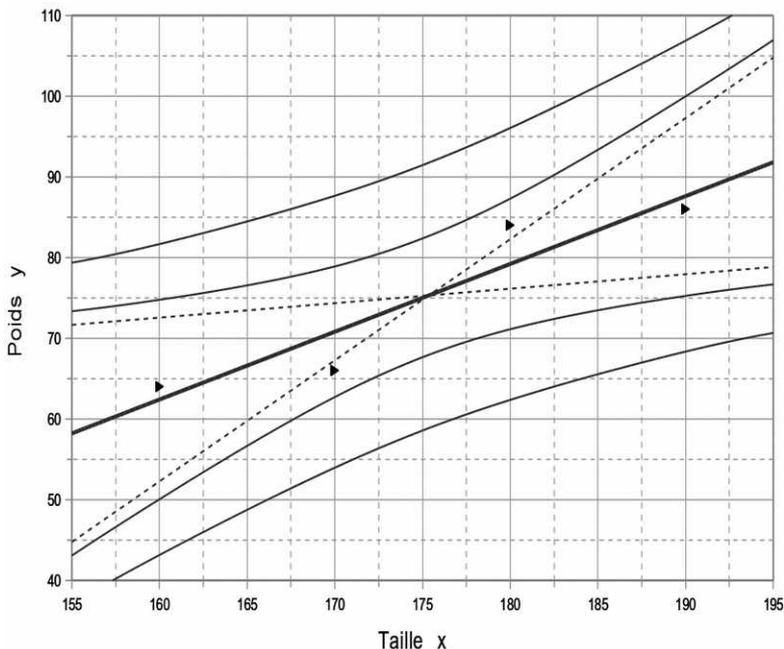
$$\Delta b \simeq 2,92 \times 39,7 \simeq 116 \quad \text{à 90\%}$$

d'où : $Poids = (0,84 \pm 0,66) Taille - (72 \pm 116)$ à 90%.

Formule ici très imprécise, ce qui n'est pas étonnant vu le peu de points et la dispersion des données. Mais la méthode de calcul est maintenant explicite et compréhensible.

Sur le graphique qui suit nous avons :

- au milieu la droite interpolée (le meilleur équilibre entre les points du dessus et ceux du dessous de cette droite).



- En pointillés sont représentés les deux droites extrêmes ($y = a_{\min} x + b_{\max}$ et $y = a_{\max} x + b_{\min}$).
- La première enveloppe correspond aux valeurs estimées de y . Intervalle de confiance de la moyenne de y_o pour une valeur x_o :

$$\Delta y_o = t_{n-2} s_r \sqrt{\frac{1}{n} + \frac{(x_o - \bar{x})^2}{\sum (x_i - \bar{x})^2}}$$

Par exemple pour $x_o = 175$ cm nous avons $y_o = 75,0 \pm 7,4$ kg. On peut même avoir une estimation en dehors de l'intervalle par exemple pour $x_o = 195$ cm nous avons $y_o = 92 \pm 15$ kg.

- La deuxième enveloppe correspond à une prédiction si nous effectuons une nouvelle mesure. Intervalle de prédiction pour une observation y_o :

$$\Delta y_o = t_{n-2} s_r \sqrt{\frac{1}{n} + \frac{(x_o - \bar{x})^2}{\sum (x_i - \bar{x})^2} + 1}$$

Par exemple, il y a 90% de chance pour une personne de 175 cm que sa masse soit entre 58 et 92 kg (en moyenne 90% des points sont dans cette seconde enveloppe et 10% en dehors).

2) Détermination du zéro absolu

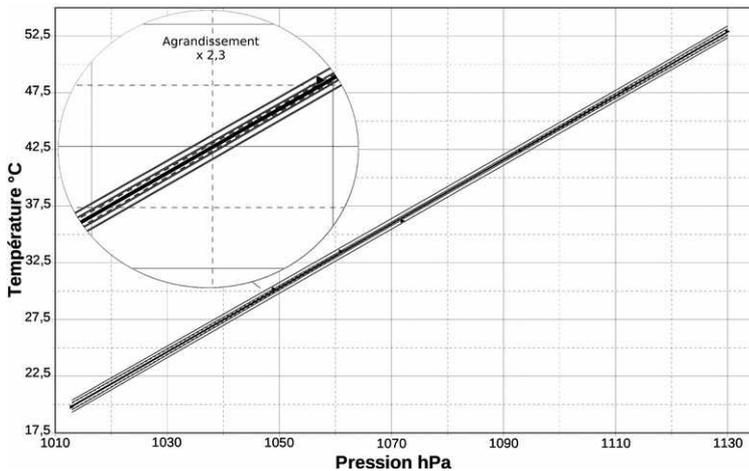
Nous étudions un gaz enfermé dans un récipient rigide de volume constant. Nous avons des sondes pour mesurer sa température et sa pression. Initialement le gaz est à température et pression ambiante, ensuite nous immergeons l'ensemble dans de l'eau chaude et nous laissons évoluer en prenant des mesures à la volée¹ :



¹ Cette expérience a réellement été réalisée le mardi 17 octobre 2006 à Bourges par M. ROUAUD et O. LEROY au Lycée Alain-Fournier.

heure	10h 15	10h 30	10h 40	10h 55	non notée	non notée	12h
température Θ (°C)	19,8	52,9	47,8	42,4	36,2	33,5	30,2
pression P (hPa)	1013	1130	1112	1093	1072	1061	1049

Nous supposons que le gaz obéit à l'équation d'état du gaz parfait $PV = nRT = nR(\Theta - \Theta_{0K})$. En traçant $\Theta(P)$ nous pouvons obtenir la température du zéro absolu : l'ordonnée à l'origine donne Θ_{0K} .



Nous voyons que la régression est bonne ($r=0,99991$) mais les points mesurés sont éloignés du zéro absolu. Par prolongement nous obtenons avec une confiance de 95% :

$$\Theta_{0K} = -266,0 \pm 4,8 \text{ °C}$$

Or, nous savons que le zéro absolu est à $-273,15\text{ °C}$, ce qui n'est pas très cohérent. Nous pouvons donc supposer qu'il y a un biais et que nous n'avons pas considéré toutes les sources d'incertitudes.

3) Régression avec barres d'erreurs

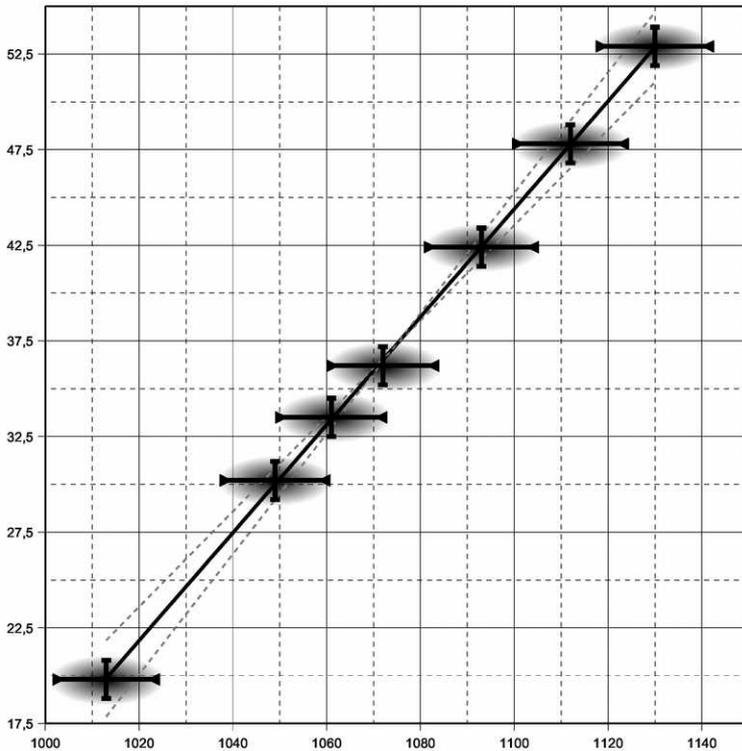
Les instruments de mesure ne sont pas parfaits et les notices indiquent leurs précisions. La précision du manomètre est à un pour cent près et celle du thermomètre au degré. Nous avons donc une incertitude sur x et y :

$$M_i(x_i \pm \Delta x_i, y_i \pm \Delta y_i)$$

Nous allons ajouter ce qu'on appelle des barres d'erreurs, nous n'avons plus des points mais des rectangles. Plus un point est précis, plus il joue un rôle important. La méthode des moindres carrés est modifiée, et la somme des carrés est pondérée par ce "poids" :

$$\sum_i w_i e_i^2 \quad \text{avec} \quad w_i = \frac{1}{(\Delta y_i)^2 + (a \Delta x_i)^2}$$

C'est une méthode itérative, nous mettons initialement une valeur de a estimée, puis la valeur de a obtenue la remplace jusqu'à l'égalité des valeurs.



Nous obtenons alors : $\Theta_{0K} = -266 \pm 35 \text{ } ^\circ\text{C}$ avec la même confiance que les incertitudes sur x_i et y_i . La valeur est, cette fois, satisfaisante. Les sources principales d'incertitudes semblent retenues.

Nous pourrions aussi considérer l'incertitude de modélisation induite par l'hypothèse de gaz parfait, mais, dans les conditions expérimentales de cette expérience, l'hypothèse est excellente. Cette source d'incertitude est négligeable devant les autres considérées ici. L'utilisation d'un modèle de gaz réel (Van der Waals par exemple) permettrait de le démontrer.

Formules [i] :

$$S^2 = \sum_i w_i [y_i - (ax_i + b)]^2$$

$$\frac{\partial S^2}{\partial b} = 0 \quad \text{et} \quad \frac{\partial S^2}{\partial a} = 0$$

conduit à :

$$b = \frac{\sum w_i y_i \sum w_i x_i^2 - \sum w_i x_i \sum w_i x_i y_i}{\Delta}$$

et

$$a = \frac{\sum w_i \sum w_i x_i y_i - \sum w_i x_i \sum w_i y_i}{\Delta}$$

avec

$$\Delta = \sum w_i \sum w_i x_i^2 - \left(\sum w_i x_i \right)^2$$

puis

$$\Delta b = \sqrt{\frac{\sum w_i x_i^2}{\Delta}}$$

et

$$\Delta a = \sqrt{\frac{\sum w_i}{\Delta}}$$

4) Linéarisation

Dans de nombreux cas nous pouvons nous ramener à un modèle linéaire (ici au sens $y = ax + b$). Nous donnons quelques exemples ci-dessous :

$y = \alpha x^\beta$ $x, y, \alpha > 0$	$y' = ax' + b$ avec $y' = \ln(y) \quad x' = \ln(x)$ et $\beta = a \quad \alpha = e^b$ et $\Delta \beta = \Delta a \quad \Delta \alpha = \alpha \Delta b$
$y = \alpha e^{\beta x} \quad x, \alpha > 0$	$y' = \ln(y) \quad x' = x$
$y = \frac{1}{\alpha + \beta x}$	$y' = \frac{1}{y}$
$y = \frac{e^{\alpha + \beta x}}{1 + e^{\alpha + \beta x}}$ (loi logis- tique)	$y' = \ln\left(\frac{y}{1-y}\right) \quad y' = \alpha + \beta x$
$y = 1 - \left(\frac{\alpha}{x}\right)^\beta$ (loi de Pareto)	$y' = \ln(1-y) \quad x' = \ln(x)$ $\beta = -a \quad \alpha = e^{-\frac{b}{a}}$
$y = 1 - e^{-\left(\frac{x}{\alpha}\right)^\beta}$ (loi de Weibull)	$y' = \ln\left(\ln\left(\frac{1}{1-y}\right)\right)$ $x' = \ln(x) \quad \beta = a \quad \alpha = e^{-\frac{b}{a}}$
$y = \alpha + e^{\beta x}$	non linéarisable
$y = \frac{\alpha x}{\beta + x}$	non linéarisable
$y = \alpha + \beta x + \gamma x^2$	non linéarisable selon $y(x)$

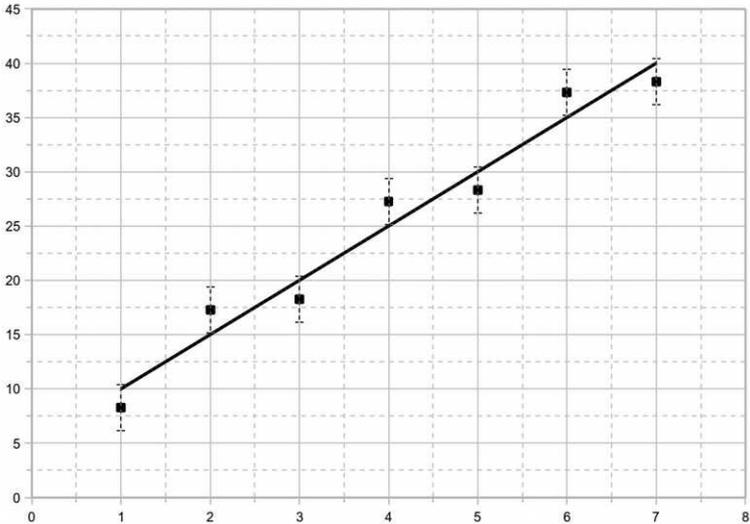
5) Comparaison des méthodes

Dans toutes les méthodes de régression que nous considérons nous nous intéressons à la variabilité de y à x fixé (la démarche inverse donnerait des résultats équivalents).

Pour un x_i donné nous avons un y_i et son écart-type σ_{y_i} .

a) Résumé

1- Régression simple



$y_1=8,3$ $y_2=17,3$ $y_3=18,3$ $y_4=27,3$ $y_5=28,3$ $y_6=37,3$ $y_7=38,3$

Cas 1 : $a=5$ $b=5$ $s_r=2,34$ $s_a=0,443$ et $s_b=1,98$

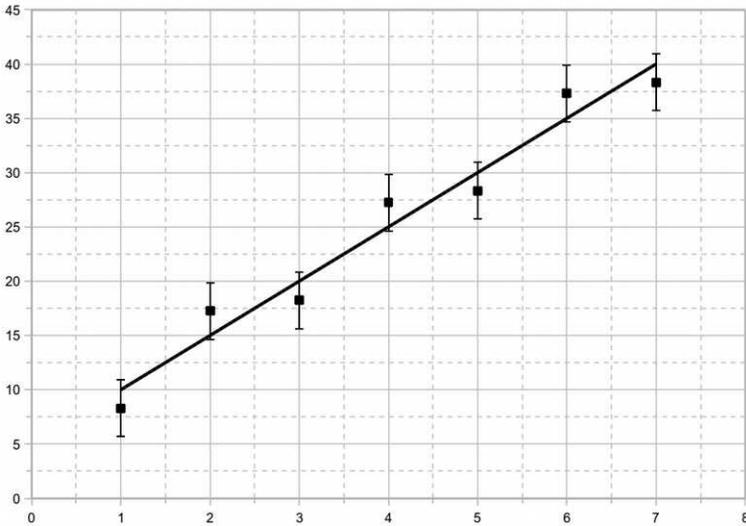
La régression simple ne correspond pas à des incertitudes sur les données nulles. Elles sont inconnues et nous les estimons à partir des données elles-mêmes. Les incertitudes

sont considérées constantes quelque soit y_i . L'incertitude correspond à l'écart-type des y_i par rapport à la droite estimée :

$$s_{y_i} = s_y = s_r = \sqrt{\frac{\sum_i (y_i - \hat{y}_i)^2}{n-2}}$$

$$s_a = \frac{s_r}{\sqrt{\sum_i (x_i - \bar{x})^2}} \quad s_b = \sqrt{\frac{\bar{x}^2}{\sum_i (x_i - \bar{x})^2}} s_r$$

2- Régression avec écarts-types constants selon y



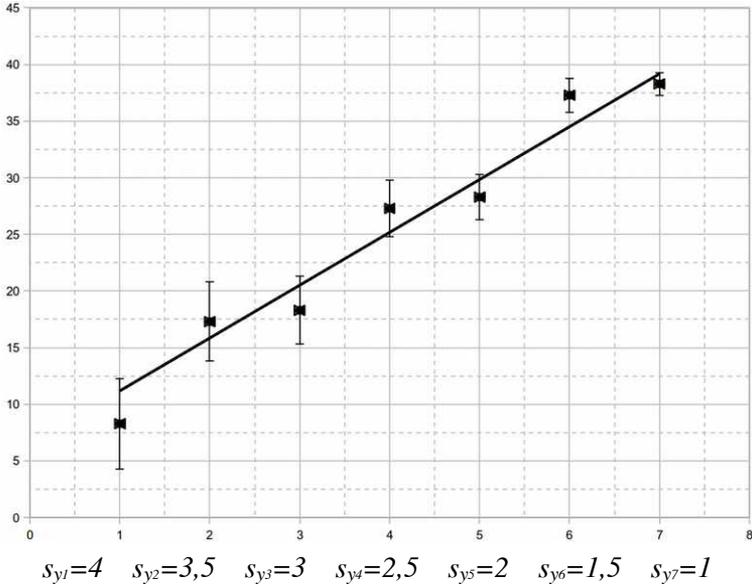
Cas 2 : $a=5$ $b=5$ $s_y=2,55$ $s_a=0,482$ $s_b=2,16$

Dans ce cas les s_{y_i} sont égaux et connus : $s_{y_i} = s_y$ et

$$s_a = \frac{s_y}{\sqrt{\sum_i (x_i - \bar{x})^2}} \quad s_b = \sqrt{\frac{\bar{x}^2}{\sum_i (x_i - \bar{x})^2}} s_y$$

Si la droite ne passe pas par les barres d'erreurs, on peut penser que toutes les sources d'incertitudes n'ont pas été retenues. Il faut, soit les intégrer dans s_y , soit appliquer la méthode précédente.

3- Régression avec écart-types selon y



Cas 3 : $a=4,67$ $b=6,51$ $s_a=0,482$ $s_b=2,16$

Les s_{y_i} sont connus. Nous pouvons appliquer la formule de propagation des écart-types :

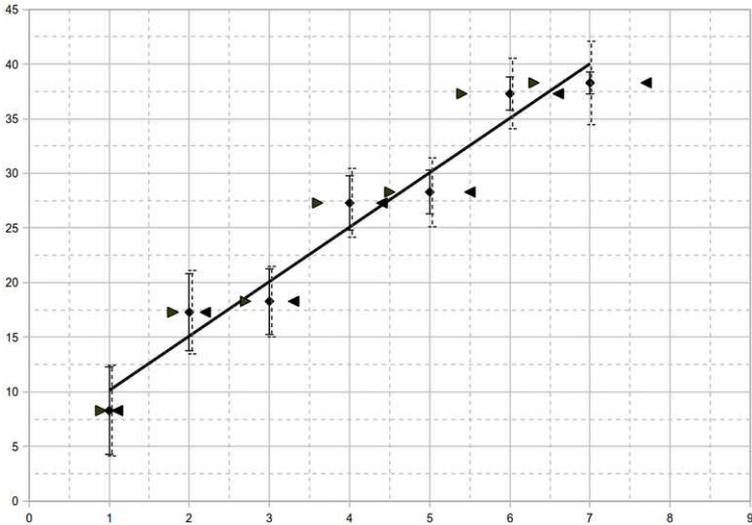
$$s_a^2 = \sum_i \left(\frac{\partial a}{\partial y_i} \right)^2 s_{y_i}^2 \quad \text{et} \quad s_b^2 = \sum_i \left(\frac{\partial b}{\partial y_i} \right)^2 s_{y_i}^2$$

Les formules sont exactes et permettent de retrouver les expressions des cas précédents. Dans le cas plus complexe

d'écart-types variables selon y_i nous avons les estimations suivantes :

$$\Delta a = \frac{1}{\sqrt{\sum w_i}} \frac{1}{\sqrt{x^2 - \bar{x}^2}} \quad \Delta b = \frac{1}{\sqrt{\sum w_i}} \sqrt{\frac{x^2}{x^2 - \bar{x}^2}} \quad w_i = \frac{1}{\sigma_{y_i}^2}$$

4- Régression avec écart-types selon y et x



$$s_{y1}=4 \quad s_{y2}=3,5 \quad s_{y3}=3 \quad s_{y4}=2,5 \quad s_{y5}=2 \quad s_{y6}=1,5 \quad s_{y7}=1$$

$$s_{x1}=0,1 \quad s_{x2}=0,2 \quad s_{x3}=0,3 \quad s_{x4}=0,4 \quad s_{x5}=0,5 \quad s_{x6}=0,6 \quad s_{x7}=0,7$$

$$s_1=4,0 \quad s_2=3,6 \quad s_3=3,4 \quad s_4=3,2 \quad s_5=3,2 \quad s_6=3,4 \quad s_7=3,6$$

$$\text{Cas 4 : } a=4,98 \quad b=5,14 \quad s_a=0,695 \quad s_b=3,16$$

Comme x_i est fixé nous reportons la dispersion de x_i

$$\text{sur } y_i : \quad s_{y_i, \text{Total}}^2 = s_i^2 = s_{y_i}^2 + a^2 s_{x_i}^2$$

À nouveau tout se passe comme si seuls les y_i avaient des

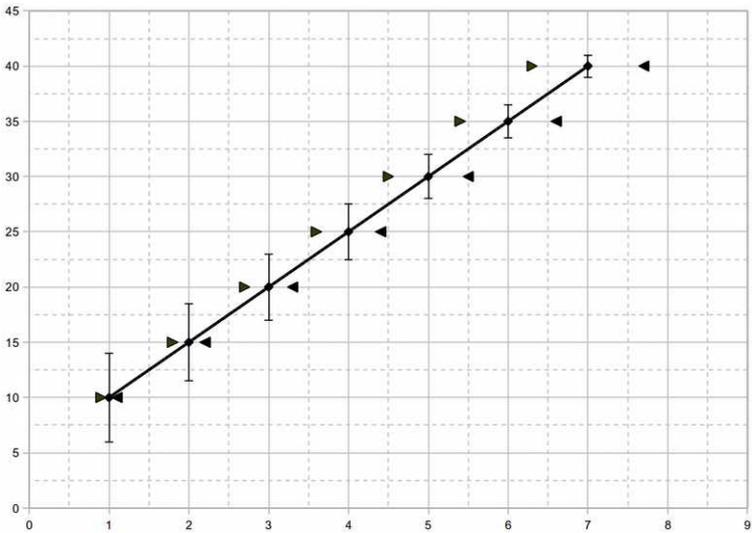
écart-types s_i . D'où les formules :

$$s_a^2 = \sum_i \left(\frac{\partial a}{\partial y_i} \right)^2 s_i^2 \quad \text{et} \quad s_b^2 = \sum_i \left(\frac{\partial b}{\partial y_i} \right)^2 s_i^2$$

Les dérivées sont complexes à calculer (les poids dépendent de a), nous pouvons par contre les évaluer numériquement. Aussi nous utilisons habituellement les estimations suivantes :

$$\Delta a = \sqrt{\frac{\sum w_i}{\Delta}} \quad \Delta b = \sqrt{\frac{\sum w_i x_i^2}{\Delta}} \quad w_i = \frac{1}{\sigma_{y_i}^2 + a^2 \sigma_{x_i}^2}$$

b) Discussion



$$\text{Cas 5 : } a=5 \quad b=5 \quad s_a=0,696 \quad s_b=3,16$$

Les incertitudes de ce cas peuvent avoir pour origine les instruments de mesure. La dispersion des valeurs du pre-

mier cas p69 peuvent provenir d'erreurs accidentelles liées à l'expérience ou d'un phénomène par essence aléatoire.

Nous aimerions que le quatrième cas p72 intègre l'ensemble les sources d'incertitudes des cas 1 et 5 :

$$s_{a_1} \simeq 0,443 \text{ et } s_{a_5} \simeq 0,696 \text{ donne } s_{a_4} \simeq 0,695$$

Nous avons utilisé les formules classiques. Nous avons l'impression que la dispersion autour de la droite de régression n'est pas incluse. Nous allons effectuer le calcul direct avec la formule de propagation et je propose la méthode des petites variations :

$$s_a^2 = \sum_i \left(\frac{\partial a}{\partial y_i} \right)^2 s_i^2 \text{ et } \left(\frac{\partial a}{\partial y_j} \right) = \lim_{\Delta y_j \rightarrow 0} \left(\frac{\Delta a}{\Delta y_j} \right)$$

En gardant tous les $y_{i \neq j}$ constants : $y_j + \Delta y_j \rightarrow a + \Delta a$

Δy_j est une petite variation par rapport à y_j , si Δy_j devient chaque fois plus petit, $\frac{\Delta a}{\Delta y_j}$ reste constant (définition de la dérivée).

Essayons de retrouver le résultat pour s_{a_5} : remplaçons $y_1 = 10$ par $y_1 = 10,001$ alors de $a = 5$ nous avons après itérations $a = 4,999907$ d'où $\frac{\Delta a}{\Delta y_1} \simeq -0,093$.

Nous remettons y_1 à 10 et répétons le procédé pour y_2 en le remplaçant par $15,001$. D'où les résultats suivants :

	$j=1$	$j=2$	$j=3$	$j=4$	$j=5$	$j=6$	$j=7$
$\frac{\partial a}{\partial y_j}$	-0,093	-0,078	-0,049	-0,006	0,041	0,080	0,105

Nous trouvons alors $s_{a_5} \simeq 0,696$, le même résultat que l'autre méthode.

Procédons de même pour s_{a_4} :

	$j=1$	$j=2$	$j=3$	$j=4$	$j=5$	$j=6$	$j=7$
$\frac{\partial a}{\partial y_j}$	-0,096	-0,080	-0,050	-0,007	0,081	0,122	0,147

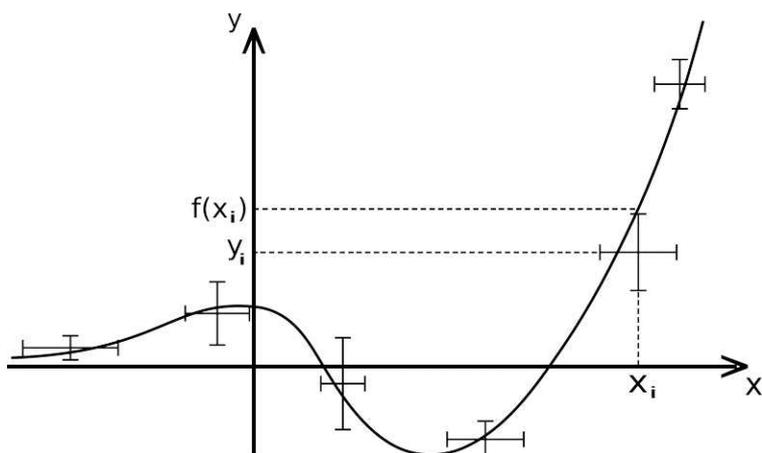
Nous trouvons alors $s_{a_4} \simeq 0,786$, résultat significativement différent de l'estimation classique qui semble maintenant intégrer l'ensemble des sources d'incertitudes.

Dans l'exercice *Écart-types proportionnel à y*, page 99, nous étudions un cas particulier où nous effectuons le calcul analytique direct. La comparaison peut ainsi être élargie.

D. Régression généralisée

Nous généralisons la méthode des moindres carrés pondérés pour une fonction et une dépendance des paramètres non linéaires. Bien que la régression multiple ait des développements analogues elle n'est pas traitée dans cet ouvrage [x].

1) Principe



Nous comparons y_i à la valeur $f(x_i)$ donnée par la fonction cherchée : $y_i - f(x_i)$. Le poids attribué à $(y_i - f(x_i))^2$ est inversement proportionnel à la variance du terme $y_i - f(x_i)$.

Les grandeurs x_i et y_i sont indépendantes¹⁴ d'où :

14 Parler d'indépendance entre deux variables alors que nous recherchons une relation fonctionnelle entre les deux peut sembler illogique. Nous faisons ici référence à la détermination expérimentale de chaque grandeur qui est indépendante au sens des incertitudes de l'autre (discuté dans le complément p179).

$$V(y_i - f(x_i)) = V(y_i) + V(f(x_i))$$

En appliquant à nouveau la formule de propagation de la variance :

$$V(f(x_i)) = f'(x_i)^2 V(x_i)$$

$$\text{dans : } S^2 = \sum_i w_i (y_i - f(x_i))^2$$

$$\text{avec } w_i = \frac{1}{\sigma_{y_i}^2 + f'(x_i)^2 \sigma_{x_i}^2} \quad 15$$

Ensuite $\frac{\partial S^2}{\partial a_k} = 0$ permet de déterminer les paramètres a_k de notre fonction (par un calcul analytique ou une résolution numérique).

À chaque fois nous pouvons nous ramener à un système d'équations linéaires de la forme $HA=B$, avec H une matrice carrée et A le vecteur associé aux paramètres. Ensuite $A=H^{-1}B$, où H^{-1} est la matrice inverse de H .

Les incertitudes sur les paramètres sont données par les termes diagonaux de la matrice H^{-1} :

$\sigma_{a_k}^2 = (H^{-1})_{kk} \sigma_r^2$ où σ_r^2 correspond à la variance des résidus par rapport à la courbe estimée.

Quand il n'y a pas d'incertitudes sur les données explicites, l'écart-type des résidus à p paramètres s'écrit :

15 S^2 est aussi appelé χ^2 . Si nous supposons les lois de probabilités normales, les écart-types peuvent être remplacés par les incertitudes en utilisant la loi de Student.

$$s_r = \sqrt{\frac{\sum (y_i - f(x_i))^2}{n - p}}$$

Quand w_i dépend de paramètres nous itérons la méthode jusqu'à ce que nous puissions considérer les poids constants. Si nous connaissons les écart-types des données, les écart-types des paramètres peuvent se calculer avec la formule de propagation, ou bien être estimés comme en régression linéaire avec barres d'erreurs p67.

2) Régression polynomiale

Dans ce cas :

$$f(x) = a_0 + a_1 x + a_2 x^2 + \dots + a_m x^m = \sum_{i=0}^m a_i x^i$$

Cette fonction est non-linéaire par rapport à x et linéaire par rapport aux paramètres.

Illustrons avec l'exemple de la loi de Cauchy due au phénomène de dispersion de la lumière :

$$n(\lambda) = a_0 + \frac{a_1}{\lambda^2} + \frac{a_2}{\lambda^4}$$

avec les données suivantes :

$\lambda(\mu m)$	0,6157	0,5892	0,5685	0,5152	0,4981
n	1,71276	1,71578	1,71852	1,72716	1,73060

Les incertitudes sur λ et n sont dans un premier temps négligées. Que valent a_0, a_1, a_2 et leurs incertitudes ?

Nous avons une régression parabolique sous la forme :

$$f(x) = a_0 + a_1 x + a_2 x^2 \quad \text{avec} \quad f = n, \quad x = 1/\lambda^2.$$

$$S^2 = \sum (y_i - a_0 - a_1 x_i - a_2 x_i^2)^2 \quad \text{et} \quad \partial S^2 / \partial a_k = 0$$

$$\text{donne :} \quad \begin{cases} \bar{y} - a_0 - a_1 \bar{x} - a_2 \bar{x}^2 = 0 \\ \overline{xy} - a_0 \bar{x} - a_1 \bar{x}^2 - a_2 \bar{x}^3 = 0 \\ \overline{x^2 y} - a_0 \bar{x}^2 - a_1 \bar{x}^3 - a_2 \bar{x}^4 = 0 \end{cases}$$

d'où $HA=B$ avec :

$$H = \begin{pmatrix} 1 & \bar{x} & \bar{x}^2 \\ \bar{x} & \bar{x}^2 & \bar{x}^3 \\ \bar{x}^2 & \bar{x}^3 & \bar{x}^4 \end{pmatrix}, \quad A = \begin{pmatrix} a_0 \\ a_1 \\ a_2 \end{pmatrix} \quad \text{et} \quad B = \begin{pmatrix} \bar{y} \\ \overline{xy} \\ \overline{x^2 y} \end{pmatrix}.$$

Avec un tableur nous pouvons facilement faire les calculs :

$$H \simeq \begin{pmatrix} 1 & 3,3 & 11 \\ 3,3 & 11 & 38 \\ 11 & 38 & 135 \end{pmatrix}, \quad H^{-1} \simeq \begin{pmatrix} 4150 & -2530 & 376 \\ -2530 & 1546 & -230 \\ 376 & -230 & 34,3 \end{pmatrix},$$

$$B \simeq \begin{pmatrix} 1,7 \\ 5,7 \\ 19 \end{pmatrix} \quad \text{d'où} \quad A = H^{-1} B \simeq \begin{pmatrix} 1,68129 \\ 0,01135 \\ 0,00022 \end{pmatrix}.$$

Pour les incertitudes :

$$\Delta a_k = \sqrt{(H^{-1})_{kk}} t_{n-3} S_r \quad \text{avec} \quad S_r = \sqrt{\frac{\sum (y_i - f(x_i))^2}{n-3}}$$

$s_r=1,87.10^{-5}$ et à 95% de confiance $t_{n-3}=4,30$

D'où :

$$a_0=1,6813\pm 0,0017, \quad a_1=(1,13\pm 0,10).10^{-2} \mu m^2$$

et $a_2=(2,2\pm 1,5).10^{-4} \mu m^4$

Prenons maintenant l'incertitude $\Delta n=0,00004$ à 95% de confiance, nous avons une incertitude sur y et pas sur x . Les poids sont donc constants : $w_i=w=1/\Delta n^2$. Nous parvenons au même système d'équations, et de même pour $H, H^{-1} B$ et A .

Ici, en procédant de manière analogue à la régression linéaire avec barres d'erreurs nous pouvons en première approche estimer la dispersion des résidus par $1/\sqrt{w_i}$:

$$\Delta a_k = \sqrt{\frac{(H^{-1})_{kk}}{\sum w_i}}, \text{ d'où :}$$

$$a_0=1,6813\pm 0,0012, \quad a_1=(1,14\pm 0,07).10^{-2} \mu m^2$$

et $a_2=(2,2\pm 1,0).10^{-4} \mu m^4$

Avec l'ensemble des incertitudes, $\Delta n=0,00004$, $\Delta \lambda=0,005 \mu m$ et $\Delta x=2\Delta \lambda/\lambda^3$, les poids dépendent du point. Nous parvenons néanmoins au même système d'équations en considérant les poids constants sur une itération, et nous calculons $H, H^{-1} B$ et A avec des paramètres estimés.

Pour calculer les moyennes nous utilisons l'expression du poids suivante :

$$w_i = \frac{1}{\Delta n^2 + (a_1 + 2a_2 x_i)^2 \Delta x_i^2}$$

Nous obtenons ainsi une première expression des paramètres. Pour itérer nous remplaçons les paramètres estimés par ces nouveaux paramètres. Nous itérons autant que nécessaire. La convergence est souvent très rapide :

Itération	a_0	a_1	a_2
Estimés	1,5	0,005	0,0001
1	1,681282848208	0,011350379724	0,000219632215
2	1,681269875466	0,011358254795	0,000218466771
...			

Ici aussi nous prenons : $\Delta a_k = \sqrt{\frac{(H^{-1})_{kk}}{\sum w_i}}$, d'où :

$$a_0 = 1,6813 \pm 0,0013, \quad a_1 = (1,14 \pm 0,08) \cdot 10^{-2} \mu m^2$$

$$\text{et } a_2 = (2,2 \pm 1,2) \cdot 10^{-4} \mu m^4$$

3) Régression non linéaire

Nous allons partir de p paramètres a_k estimés et utiliser une méthode d'itération. Sur chaque itération les poids seront considérés constants et la fonction sera linéarisée pour chacun des n points sur l'ensemble des paramètres.

La fonction dépend des x_i et des paramètres a_k .

Nous posons : $f_i = f(x_i; a_k)$.

$$S^2 = \sum_i w_i (y_i - f_i)^2 \quad \text{et} \quad \frac{\partial S^2}{\partial a_k} = -2 \sum_{i=1}^n w_i \frac{\partial f}{\partial a_k} (y_i - f_i) = 0$$

Les premiers paramètres estimés sont notés a_{0k} . ensuite nous les noterons a_{jk} pour la $j^{\text{ème}}$ itération. Nous allons effectuer un développement limité d'ordre un pour une petite variation δa_{0k} autour de a_{0k} ¹⁶.

Notons $\vec{a} = (a_1, \dots, a_k, \dots, a_p)$ et $\vec{\delta a} = (\delta a_1, \dots, \delta a_k, \dots, \delta a_p)$:

$$f(x_i; \vec{a}_0 + \vec{\delta a}_0) = f(x_i; \vec{a}_0) + \sum_{l=1}^p \delta a_{0l} \left(\frac{\partial f(x_i; \vec{a})}{\partial a_l} \right)_{a_{0l}}$$

soit en allégeant les notations :

$$f_i = f_{0,i} + \sum_l \left(\frac{\partial f}{\partial a_l} \right)_{0,i} \delta a_{0l}$$

$$\text{ainsi} \quad \sum_i w_i \frac{\partial f}{\partial a_k} (y_i - f_{0,i} - \sum_l \left(\frac{\partial f}{\partial a_l} \right)_{0,i} \delta a_{0l}) = 0$$

$$\text{et} \quad \sum_i w_i \frac{\partial f}{\partial a_k} (y_i - f_{0,i}) = \sum_{i,l} w_i \frac{\partial f}{\partial a_k} \frac{\partial f}{\partial a_l} \delta a_{0l}.$$

$$\text{Nous posons} \quad H_{k,l} = \sum_i w_i \left(\frac{\partial f}{\partial a_k} \right)_i \left(\frac{\partial f}{\partial a_l} \right)_i = H_{l,k},$$

$$B_k = \sum_i w_i \frac{\partial f}{\partial a_k} (y_i - f_{0,i}) \quad \text{et} \quad A_l = \delta a_{0l}.$$

d'où à nouveau $HA=B$ et $A=H^{-1}B$. Nous itérons jusqu'à convergence et variations sur les paramètres négligeables.

16 MATH : nous généralisons la notion de dérivée en ajoutant les variations selon tous les paramètres :

$$f(x_0 + \epsilon) \simeq f(x_0) + \epsilon (f'(x))_{x_0}$$

Ici aussi nous utilisons : $\sigma_{\alpha_k}^2 = (H^{-1})_{kk} \sigma_r^2$.

Illustrons avec une expérience de biologie, où l'on étudie la relation entre la concentration du substrat $[S]$ et la vitesse de réaction dans une réaction enzymatique à partir de données reportées dans le tableau suivant [ix] :

i	1	2	3	4	5	6	7
$[S]$	0,038	0,194	0,425	0,626	1,253	2,500	3,740
v	0,050	0,127	0,094	0,2122	0,2729	0,2665	0,3317

Nous souhaitons ajuster sous la forme :

$$y = \frac{\alpha x}{\beta + x} \quad \text{en posant } v=y \text{ et } [S]=x.$$

Nous partons des estimations $\alpha_0=0,9$ et $\beta_0=0,2$.

$$\frac{\partial f}{\partial \alpha} = \frac{x}{\beta + x}, \quad \frac{\partial f}{\partial \beta} = -\frac{\alpha x}{(\beta + x)^2}, \quad H = \begin{pmatrix} H_{11} & H_{12} \\ H_{21} & H_{22} \end{pmatrix},$$

$$A = \begin{pmatrix} \alpha \\ \beta \end{pmatrix} \quad \text{et} \quad B = \begin{pmatrix} B_1 \\ B_2 \end{pmatrix}.$$

En absence d'incertitudes fournies sur x_i et y_i :

$$H_{11} = \sum_{i=1}^7 \left(\frac{\partial f}{\partial \alpha} \right)_{\alpha_0, i} \left(\frac{\partial f}{\partial \alpha} \right)_{\alpha_0, i} = \sum_i \left(\frac{\partial f}{\partial \alpha} \right)_{\alpha_0, i}^2 = \sum_i \left(\frac{x_i}{\beta_0 + x_i} \right)^2,$$

$$H_{12}=H_{21}=\sum_i \frac{\alpha_0 x_i^2}{(\beta_0+x_i)^3} \text{ et } H_{22}=\sum_i \frac{\alpha_0^2 x_i^2}{(\beta_0+x_i)^4} .$$

$$B_1=\sum_i \frac{x_i}{\beta_0+x_i} (y_i - f_{0,i}) \text{ et } B_2=-\sum_i \frac{\alpha_0 x_i}{(\beta_0+x_i)^2} (y_i - f_{0,i})$$

$$\text{avec } f_{0,i}=\frac{\alpha_0 x_i}{\beta_0+x_i}$$

Nous pouvons maintenant tout mettre dans un tableur, nous obtenons :

$$H_{11}\simeq 3,81, \quad H_{12}=H_{21}\simeq -2,89, \quad H_{22}\simeq 3,70,$$

$$B_1\simeq -2,33 \text{ et } B_2\simeq 1,86 .$$

puis

$$H^{-1}\simeq \begin{pmatrix} 0,649 & 0,508 \\ 0,508 & 0,668 \end{pmatrix} \text{ d'où } A=H^{-1}B\simeq \begin{pmatrix} -0,567 \\ 0,0602 \end{pmatrix} .$$

Ainsi de $\alpha_1=\alpha_0+\delta\alpha_0$ et $\beta_1=\beta_0+\delta\beta_0$ nous avons les nouveaux paramètres estimés :

$$\alpha_1\simeq 0,9-0,567\simeq 0,333 \text{ et } \beta_1\simeq 0,2+0,06\simeq 0,260$$

Nous remplaçons ces valeurs au début des calculs à la place de α_0 et β_0 pour itérer. Nous calculons de nouvelles matrices et vecteurs H, H^1, B et A , et nous obtenons α_2 et β_2 .

Nous donnons l'ensemble des résultats dans le tableau suivant :

Itérat°	α	β	$\delta\alpha$	$\delta\beta$	S^2
	0,9	0,2	-0,57	0,060	1,4454965815
1	0,333	0,26	0,0101	0,166	0,0150720753
2	0,343	0,43	0,0150	0,103	0,0084583228
3	0,358	0,53	0,0040	0,024	0,0078643240
4	0,3614	0,554	0,0004	0,0024	0,0078441826
5	0,36180	0,5561	0,00003	0,00018	0,0078440067
6	0,36183	0,5563	0,000002	0,000013	0,0078440057

Après suffisamment d'itérations : $H^{-1} \simeq \begin{pmatrix} 1,52 & 6,34 \\ 6,34 & 36,2 \end{pmatrix}$

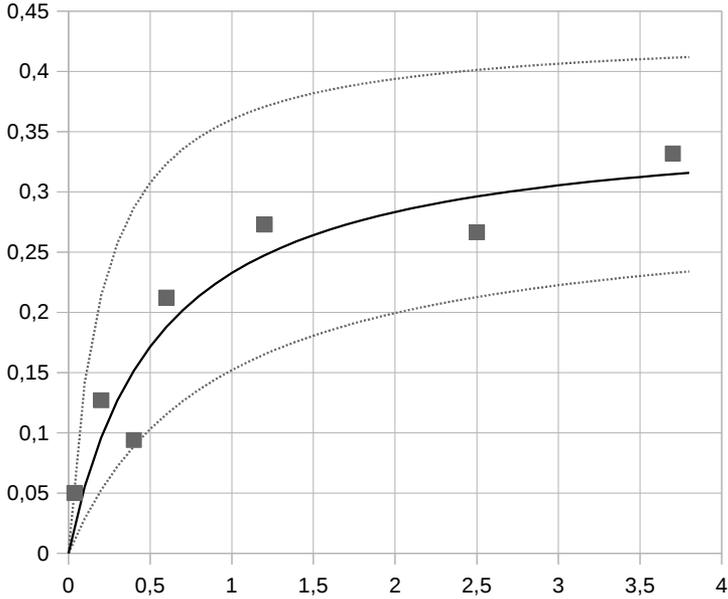
Calculons les incertitudes sur les paramètres :

$$s_r = \sqrt{\frac{S^2}{n-2}} \text{ et } \Delta \alpha = \sqrt{(H^{-1})_{11}} t_{n-2} s_r \simeq \sqrt{1,52} \cdot 1,48 \cdot \sqrt{\frac{0,00784}{5}}$$

$$\text{d'où } \Delta \alpha \simeq 0,07, \quad \Delta \beta \simeq 0,35.$$

Finalement : $\alpha = 0,36 \pm 0,07$ et $\beta = 0,56 \pm 0,35$ à 80%.

Le graphique qui suit représente la vitesse de réaction en fonction de la concentration du substrat. Les carrés sont les points expérimentaux, la ligne continue la courbe optimale et en pointillés les deux courbes extrêmes $f_{\alpha_{\max}, \beta_{\min}}$ et $f_{\alpha_{\min}, \beta_{\max}}$.



Si nous voulons considérer les incertitudes pour chaque point nous devons ajouter les poids. Ceux-ci sont considérés constants sur chaque itération. Il faudra calculer la dérivée de f par rapport à x qui intervient dans l'expression du poids.

Pour des écart-types sur l'ensemble des données nous pouvons calculer les écart-types des paramètres en s'inspirant des méthodes de la régression linéaire p69.

E. Exercices

Exercice 1 : Corrélations *corrigé en version complète*

1- On réalise neuf expériences, nous obtenons à chaque fois trois réalisations des grandeurs X_1 , X_2 et X_3 :

$i=$	1	2	3	4	5	6	7	8	9
X_1	$x_1^1 = -1$	-1	-1	0	0	0	1	1	1
X_2	$x_2^1 = 1$	0	-1	1	0	-1	1	0	-1
X_3	$x_3^1 = -1$	0	1	-1	0	1	-1	0	1

a) Sur l'ensemble des expériences déterminez les moyennes arithmétiques et les écart-types de ces trois grandeurs.

b) Tracez X_2 en fonction de X_1 . Puis $X_3(X_1)$ et $X_3(X_2)$.

c) Calculez les coefficients de corrélation r_{12} , r_{13} et r_{23} . Commentez.

2- Même chose avec les données suivantes :

X_1	0	1	-1	2	0	-1	0	-2	1
X_2	1	2	-1	2	0	-2	-1	-2	1

3- Même chose avec les données suivantes :

X_1	-1	2	-2	0	-2	2	1
X_2	-1	0	2	-2	0	2	-1

Exercice 2 : Volumes

corrigé en version complète

À l'aide d'une pipette jaugée nous remplissons quatre béchers avec 100 mL d'eau chacun. Pour tester la pipette et connaître précisément la quantité d'eau, nous effectuons quatre pesées au décigramme et nous obtenons, ramené en mL, les résultats suivants pour les différents béchers :

$$V_1 = \{100,1 ; 100,0 ; 99,9 ; 100,0\}$$

1- Calculez la moyenne et l'écart-type de V_1 . Estimez la précision de la pipette avec une confiance de 95%.

Nous remplissons maintenant deux béchers et rassemblons le contenu des deux dans un seul :

$$V = V_1 + V_2.$$

Correspondant à V_1 , nous avons les mesures suivantes pour V_2 :

$$V_2 = \{100,0 ; 100,1 ; 100,0 ; 99,9\}$$

Par exemple pour la troisième mesure nous avons $V_1 = 99,9$ mL et $V_2 = 100,0$ mL.

2- Montrez que V_1 et V_2 sont des grandeurs indépendantes.

3- Calculez la moyenne de V , son écart-type et l'incertitude ΔV à 95%.

4- Pourriez-vous retrouver ce résultat avec la formule de propagation des incertitudes?

(Pour affiner le test il faudrait prendre plus de mesures, mais le principe reste le même, et les résultats restent valides car nous avons élargi avec le Student, considéré des données décorréelées et des paquets globalement gaussiens. Nous devrions aussi tenir compte des incertitudes sur les mesures -résolution- en plus de leur dispersion.)

Exercice 3 : Arbres *corrigé en version complète*

Nous voulons mesurer la distance d entre deux arbres. Pour cela nous disposons d'un bâton d'une longueur d'un mètre. D'un arbre à l'autre nous reportons le bâton cent fois. Nous estimons pour chaque report une incertitude de 1 cm.

Quelle est l'incertitude sur la valeur de d ?

Exercice 4 : Méthode de Bessel
corrigé en version complète

C'est une méthode de focométrie qui permet de mesurer la distance focale f d'une lentille convergente. Pour cela on mesure la distance D entre un objet lumineux et son image sur un écran. Quand $D > 4f$ il existe deux positions où l'image est nette. La distance entre ces deux positions est notée d . Nous avons ensuite la distance focale de la lentille par la relation $f = (D^2 - d^2) / 4D$. Nous mesurons $D = 2000 \pm 10$ mm et $d = 536 \pm 20$ mm.

Quelle est alors l'incertitude sur f ?

Exercice 5 : Indice *corrigé en version complète*

Nous voulons déterminer expérimentalement l'indice n_2 d'un verre. Pour cela nous effectuons l'expérience de la réfraction d'un faisceau laser. Les lois de Descartes pour l'optique géométrique indique que $n_1 \cdot \sin(i_1) = n_2 \cdot \sin(i_2)$, avec n_1 les indices lumineux et i_1 les angles par rapport à la normale au dioptre. Nous avons $n_1 = n_{\text{air}} = 1$, $i_1 = 30 \pm 1^\circ$ et $i_2 = 20 \pm 2^\circ$.

Déterminez n_2 avec son incertitude.

Exercice 6 : Formule de Cauchy
corrigé en version complète

Nous voulons mesurer la variation d'indice lumineux n en fonction de la longueur d'onde λ dans un milieu transparent (phénomène de dispersion). Pour cela nous utilisons un prisme, une lampe à sodium et un goniomètre. D'après la théorie la variation de $n(\lambda)$ doit suivre la formule de Cauchy dans le spectre visible :

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

Le spectre de raie du sodium est connu. Pour chaque raie de longueur d'onde λ_i on calcule l'indice n_i correspondant en utilisant la formule du prisme suivante :

$$n_i = \frac{\sin\left(\frac{A + D_{m,i}}{2}\right)}{\sin\left(\frac{A}{2}\right)}$$

D_m est la déviation minimale. $A=60^\circ$ est l'angle au sommet du prisme. Ces deux angles sont mesurés à 2' près (1' = une minute d'arc et $1^\circ = 60'$).

Nous obtenons ainsi le tableau de données suivant :

$\lambda(nm)$	615,7	589,2	568,5	515,2	498,1
Couleur	rouge	jaune	vert-jaune	vert	bleu-vert
D_m	57° 49,5'	58° 9'	58° 28'	59° 26,5'	59° 50'
n	1,71276	1,71568	1,71852	1,72716	1,73060

1- Déterminez l'incertitude sur n (celle-ci est globalement constante d'une valeur à l'autre).

2- Acceptant l'expression en $1/\lambda^2$ de la formule de Cauchy, en déduire A et B avec leurs incertitudes. Que vaut le coefficient de régression r ?

3- Peut-être qu'en traçant n en fonction de $1/\lambda$ ou $1/\lambda^3$ nous aurions un meilleur alignement des points ?

Nous voulons vérifier expérimentalement que la variation en $1/\lambda^2$ est bien la meilleure des relations polynomiales. Pour cela nous prenons la forme :

$$n(\lambda) = A + B \cdot \lambda^\alpha$$

Proposez une méthode pour déterminer A , B et α .

Nous pourrions vérifier le modèle car nous aurons α avec son incertitude.

Exercice 7 : Mur *corrigé en version complète*

Nous avons un mur d'une surface $S=72 \text{ m}^2$. La température extérieure est de 6°C et la température intérieure est maintenue à 18°C . Ce mur de 50 cm d'épaisseur est constitué de $e_p=40 \text{ cm}$ de paille compressée (conductivité thermique $\lambda_p=45 \text{ mW/K/m}$) et de $e_e=10 \text{ cm}$ d'enduit ($\lambda_e=200 \text{ mW/K/m}$). Les λ sont données à 10% près, les épaisseurs au cm et les température au demi degré.

1- Déterminez la résistance thermique, avec son incertitude, de la paille pour ce mur ($R \cdot \lambda \cdot S = e$)

2- Même chose pour l'enduit.

3- Sachant que les résistances thermiques s'associent comme les résistances électriques en série, déterminez la résistance thermique totale du mur avec son incertitude.

4- Quelle doit être la puissance minimale du chauffage de la maison rien que pour compenser les pertes par les murs ? ($\Delta T = R \cdot \Phi$)

Exercice 8 : Isolation et inertie

corrigé en version complète

Nous avons une maison proche des préconisations pour une *maison énergie zéro*. Les résistances thermiques ramenées au m^2 , soit e/λ , sont pour le toit de $8 m^2.K/W$ pour les murs et le sol de $4 m^2.K/W$ et pour les huisseries de $1 m^2.K/W$ (les résistances sont connues avec une précision de 10%). La maison a une surface au sol de $36 m^2$, $54 m^2$ de toit, $82 m^2$ de mur et $8 m^2$ d'huisseries.

1- Les résistances équivalentes du toit, du mur, du sol et des huisseries étant en parallèle, déterminez la résistance thermique totale (en K/W) avec sa précision.

Les températures extérieures et intérieures sont constantes.

2- Quelle doit être la puissance minimale du chauffage de la maison pour maintenir la température intérieure constante en compensant les pertes ?

3- Nous coupons le chauffage et nous mesurons la température au cours du temps :

t en heures	0	1	2	4	5	6	8	9	10
T en °C	18	16	14	12	11	10	9	9	8

Expliquez pourquoi la décroissance ne peut être linéaire. Nous considérons la décroissance de type exponentielle : $T(t)=a.exp(-t/\tau)+b$. Déterminez b , a et τ , avec leurs incertitudes.

4- La maison est isolée par l'extérieur. Le flux thermique perdu correspond à une diminution de l'énergie emmagasinée dans la maison. Cette inertie est due à la capacité thermique C des matériaux (J/K).

a) En raisonnant sur un intervalle de temps infinitésimal dt trouvez l'équation différentielle

vérifiée par $T(t)$ et retrouvez l'expression du 3.

b) Quelle est la relation entre τ , R et C ? Déterminez C et son incertitude.

Au 3. nous pouvons aussi tenir compte des incertitudes de mesure : le temps peut-être considéré comme parfaitement connu et la température est mesuré avec un thermomètre à dilatation d'un liquide avec des graduations tous les degrés Celsius.

Pour simplifier nous avons considéré que la température extérieure reste constante (pour tenir compte des variations jour/nuit nous envisagerions des variations sinusoïdales et une approche harmonique).

Exercice 9 : Rendement corrigé en version complète

Des quantités d'engrais déterminées sont répandues sur des champs et nous obtenons les rendements suivants :

Engrais (kg/ha)	100	200	300	400	500	600	700
Rendement (Q/ha)	41	44	53	63	66	65	78

1- Déterminez la droite de régression qui passe par ce nuage de points. Pente, ordonnée à l'origine et incertitudes avec une confiance de 95%.

2- Pour 550 kg d'engrais par ha estimez le rendement.

3- Même chose en absence d'engrais.

4- Si un agriculteur répand 250 kg d'engrais par hectare, quelle est la probabilité qu'il obtienne 40 à 60 quintaux de céréales ?

Exercice 10 : Étude d'une pile
corrigé en version complète

Nous cherchons à déterminer la tension à vide E et la résistance interne r .

Pour cela nous mesurons pour la pile différentes valeurs de U et I avec un voltmètre et un ampèremètre ($U = E - r.I$) :

calibre pour U :	unité : V précision $\pm 0,05\% \pm 0,003$										
U (V)	4,731	4,731	4,730	4,728	4,724	4,724	4,722	4,721	4,719	4,716	
calibres pour I :	unité : μA précision $\pm 0,2\% \pm 0,03$			unité : mA précision $\pm 0,2\% \pm 0,0003$							
I	92,83	115,45	152,65	0,2352	0,4686	0,5200	0,5841	0,6661	0,7750	0,9264	

1- Sans les barres d'erreurs : déterminez $E \pm \Delta E$ et $r \pm \Delta r$.

2- Même chose en incluant cette fois les incertitudes de chaque mesure indiquées dans la notice du fabricant du multimètre.

Exercice 11 : Focométrie *corrigé en version complète*

Nous désirons déterminer la distance focale f' d'une lentille convergente. La lentille est en O et fait l'image net en A' d'un point objet A . Nous mesurons OA , OA' et leurs incertitudes (sont incluses toute les sources d'incertitudes, géométriques, latitude de mise au point et modélisation). Nous considérons que la lentille vérifie la relation de conjugaison :

$$1/OA' + 1/OA = 1/f'$$

Déterminez f' par l'intermédiaire d'une régression linéaire. Tableau page suivante.

Données expérimentales (mm) :			
OA	ΔOA	OA'	$\Delta OA'$
635	5	150	15
530	5	160	17
496	5	164	15
440	5	172	18
350	5	191	20
280	5	214	25
210	5	292	28
150	5	730	102

Théorie

Exercice 12 : *corrigé en version complète*

Pour la régression simple, montrez que :

$$\sum_{i=1}^n (x_i - \bar{x})^2 = n(\overline{x^2} - \bar{x}^2)$$

Exercice 13 : *corrigé en version complète*

Pour la régression simple, montrez que nous pouvons

aussi écrire $s_b = s_r \sqrt{\frac{1}{n} + \frac{\bar{x}^2}{\sum (x_i - \bar{x})^2}}$.

Exercice 14 : *corrigé en version complète*

Pour la régression simple, montrez que : $\Delta b = \sqrt{x^2} \Delta a$.

Exercice 15 : Asymptotes *corrigé en version complète*

Pour la régression simple, montrez que au dessus et en dessous de la droite de régression les asymptotes des droites extrêmes, des courbes de confiance et de prédiction se confondent quand x_o devient grand.

Exercice 16 : Intervalle de confiance et de prédiction pour la régression linéaire avec barres d'erreurs *corrigé en version complète*

Régression avec barres d'erreurs

1- Donnez les expressions de \bar{x} et \bar{x}^2 en utilisant les poids w_i . Quel lien constatez-vous avec l'expression de Δ (p67) ?

Analogie

2- À partir des intervalles de confiance et de prédiction pour y_o en régression simple (p62 et suivante) déterminez par analogie les formules qui suivent :

estimation :

prédiction :

$$\Delta y_o = \frac{1}{\sqrt{n \sum w_i}} \sqrt{1 + \frac{(x_o - \bar{x})^2}{x^2 - \bar{x}^2}} \quad \Delta y_o = \frac{1}{\sqrt{n \sum w_i}} \sqrt{1 + n + \frac{(x_o - \bar{x})^2}{x^2 - \bar{x}^2}}$$

3- Déterminez les écarts selon y à la droite de régression pour les droites extrêmes, les courbes de confiance et les courbes de prédiction quand x_o devient grand.

Dans un exercice précédent nous avons montré que pour la régression linéaire les asymptotes se confondaient à l'infini. Par analogie que doit-on poser pour qu'il en soit de même en régression avec barres d'erreurs ?

Montrer que nous avons alors les formules suivantes :

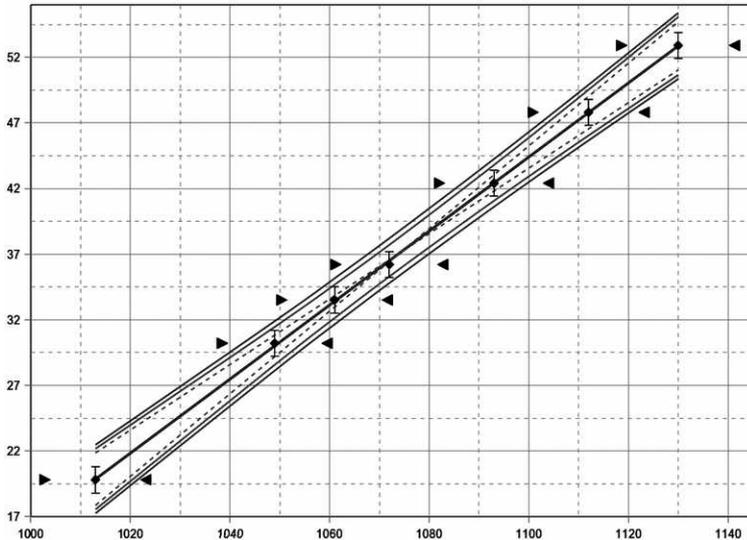
Estimation :

Prédiction :

$$\Delta y_o = \frac{1}{\sqrt{\sum w_i}} \sqrt{1 + \frac{(x_o - \bar{x})^2}{x^2 - \bar{x}^2}} \quad \Delta y_o = \frac{1}{\sqrt{\sum w_i}} \sqrt{2 + \frac{(x_o - \bar{x})^2}{x^2 - \bar{x}^2}}$$

Les formules ainsi obtenues par analogie sont empiriques, elle semblent expérimentalement cohérentes, et demandent confirmation par une démonstration théorique.

Courbes de confiance et prédiction pour l'expérience du zéro absolu : T(°C) en fonction de P(hPa)



Exercice 17 : Autres expressions

corrigé en version complète

Pour la régression avec barres d'erreurs donnez l'expression de a , b , Δa et Δb en fonction de \bar{x} , \bar{y} , \overline{xy} , $\overline{x^2}$ et $\overline{y^2}$. Comparez avec la régression simple.

Exercice 18 : Méthode des moindres carrés

corrigé en version complète

Démontrez par la méthode des moindres carrés les expressions de a et b :

1- pour la régression linéaire simple.

2- pour la régression linéaire avec barres d'erreurs.

Les Δx_i et Δy_i sont considérés petits par rapport à x_i et y_i .

Démonstration des expressions de Δa et Δb pour la régression simple :

Méthode 1 :

1- Montrez que $a = \sum_i p_i y_i$ avec $p_i = \frac{x_i - \bar{x}}{\sum (x_i - \bar{x})^2}$.

2- Dédurre de cette formule de a sa variance $V(a)$ ¹⁷.

Méthode 2 :

Utilisez la formule de propagation des écart-types.

Pour la régression linéaire simple, pourriez-vous retrouver a , b , Δa et Δb en utilisant la méthode matricielle de la régression généralisée ?

Exercice 19 : Espérance de a

corrigé en version complète

Pour la régression linéaire, nous notons α et β les paramètres de la population : $E(y_i) = \alpha x_i + \beta$.

a et b sont les paramètres estimés à partir d'un échantillon : $\hat{y}_i = a x_i + b$

Montrez que nous avons un estimateur non biaisé pour α , soit $E(a) = \alpha$.

¹⁷ MATH : $E(X+Y) = E(X) + E(Y)$. Si X et Y sont deux variables indépendantes : $V(X+Y) = V(X) + V(Y)$.

Exercice 20 : Écart-types proportionnels à y
corrigé en version complète

Nous nous plaçons dans le cas théorique particulier où les écart-types de la régression linéaire sont proportionnels à y : $\sigma_{y_i} = k y_i$. Ce cas, où les incertitudes relatives sont constantes, est expérimentalement courant. Nous nous intéressons ici au coefficient directeur a .

1- Montrez que :
$$a = \frac{\sum \frac{1}{y^2} \sum \frac{x}{y} - \sum \frac{x}{y^2} \sum \frac{1}{y}}{\sum \frac{1}{y^2} \sum \frac{x^2}{y^2} - \left(\sum \frac{x}{y^2} \right)^2}$$

2- Exprimez $\frac{\partial a}{\partial y_j}$ (calcul long).

3- Calculez s_a en utilisant les expressions trouvées pour les deux jeux de données suivants :

x_i	1	2	3	4	5	6	7
1 : y_i	10	15	20	25	30	35	40
2 : y_i	8,286	17,286	18,286	27,286	28,286	37,286	38,286

(Nous prenons $k=0,1$).

Retrouver ces résultats avec la méthode numérique (évaluation des dérivées par petites variations).

Comparer avec les valeurs obtenues par la méthode classique.

Exercice 21 : Interprétation de l'expression de w_i
corrigé en version complète

Justifiez graphiquement la position de a dans l'expression de w_i .

Régression non-linéaire

Exercice 22 : Décomposition en gaussiennes

corrigé en version complète

Une fabrique usine des clous. Nous mesurons avec un pied à coulisse la taille de 48 clous :

Taille	59,97	59,98	59,99	60,00	60,01	60,02	60,03	60,04	60,05
Nombre	2	4	6	5	5	10	9	6	1

Deux machines fabriquent les clous. Les clous fabriqués par une machine n'ont pas exactement la même taille, nous supposons que la taille est distribuée selon un profil gaussien. Déterminer les caractéristiques des profils gaussiens des clous fabriqués par chacune des deux machines (maximum, moyenne et écart-type). Combien de clous ont fabriqués chaque machine ?

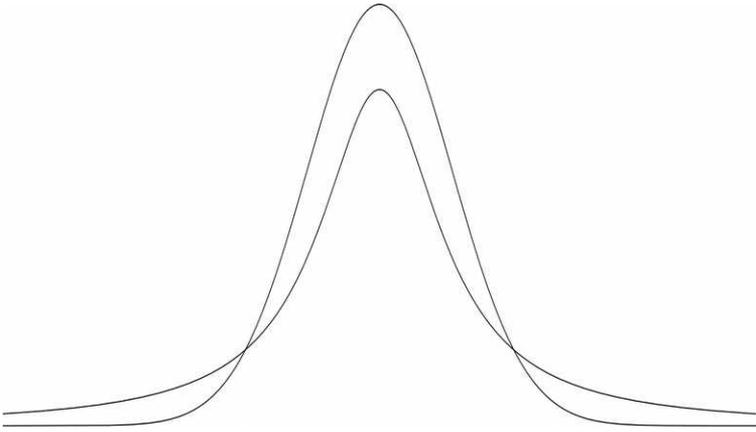
Pour vérifier vos calculs de corrélation vous pouvez utiliser la fonction du tableur OOo :

COEFFICIENT.CORRELATION(*,**).

Courbes : Insertion>Diagramme...>XY(Dispersion)>etc.

Pour les calculs matriciels, inversion de matrices : INVERSEMAT(*), produit de matrices : PRODUITMAT(*,**).

Vous pouvez utiliser le fichier IncertitudesLibresOOo32.ods sur le site www.incertitudes.fr pour réaliser les régressions. Feuille 2, régression simple et feuille 3 avec barres d'erreurs.



III. LOIS DE PROBABILITÉS

Nous allons lister différentes lois de probabilités discrètes puis continues. Nous montrerons qu'il s'agit bien de lois de probabilité et nous déterminerons leur espérance $\mu = E(X)$ et variance $V = E[(X - \mu)^2]$. La variance pourra aussi se calculer avec la formule $V = E[X^2] - E[X]^2$ et nous aurons ainsi l'écart-type $\sigma = \sqrt{V}$.

Nous pourrions calculer d'autres moments qui permettent de déterminer par exemple la symétrie et l'aplatissement de la distribution de probabilité.

Les moments ordinaires : $\mu_k = E(X^k)$

Les moments centrés réduits sont instructifs car caractéristiques de la forme de la distribution :

$$\beta_{k-2} = E\left[\left(\frac{X - \mu}{\sigma}\right)^k\right] \quad \text{ou} \quad \beta_{k-2} = \frac{\mu_k}{\sigma^k}$$

β_1 : coefficient d'asymétrie (moment centré réduit d'ordre trois)

β_2 : coefficient d'aplatissement

Nous nous intéresserons aussi à la somme de variables aléatoires indépendantes : $Z = X + Y$.

A. Lois discrètes

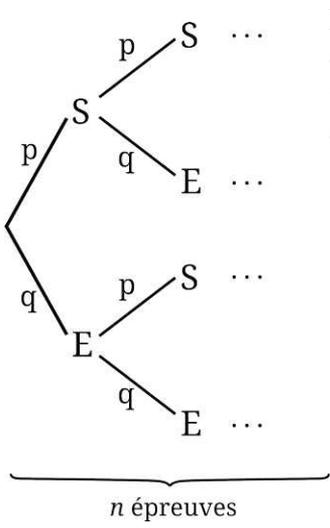
1) Loi binomiale

Nous considérons une succession de n épreuves identiques et indépendantes. Chaque épreuve a deux issues que nous nommons succès et échec. La probabilité du succès est notée $p=P(S)$. Les deux paramètres de la loi sont n et p et la loi peut s'écrire $\mathcal{B}(n,p)$.

Nous voulons déterminer la probabilité d'avoir k succès suite aux n épreuves. Un chemin comportant k succès comporte nécessairement $n-k$ échecs et sa probabilité vaut $p^k q^{n-k}$, avec $q=1-p$, probabilité de l'échec.

Il suffit ensuite de dénombrer le nombre de chemins qui comprennent k succès, manières de choisir k objets parmi n . n choix pour la position du premier succès, $n-1$ pour le deuxième et $n+1-k$ pour le k^{em} succès :

$n(n-1)\dots(n+1-k)=n!/(n-k)!$ possibilités¹⁸.



Nous divisons ensuite par $k!$ pour enlever les comptages multiples (par exemple S_1S_2E et S_2S_1E correspondent au même chemin).

D'où :

$$P(X=k) = \binom{n}{k} p^k q^{n-k}$$

$$\text{avec } \binom{n}{k} = \frac{n!}{(n-k)!k!}$$

(se dit « k parmi n »)

$$\text{et } \sum_{k=0}^n P(X=k) = 1$$

¹⁸ $n!$, se lit « n factorielle » avec $n!=1 \times 2 \times 3 \dots \times n$.

On montre que $E(X)=np$ et $V(X)=npq$.

La somme de deux lois binomiales indépendantes de même p et aussi une loi binomiale. La somme d'une $\mathcal{B}_x(n_1, p)$ et d'une $\mathcal{B}_y(n_2, p)$ est une $\mathcal{B}_z(n_1+n_2, p)$.

Pour déterminer la loi de Z somme de X et Y nous utilisons, pour les lois discrètes, la propriété suivante :

$$P(Z=k)=\sum_i P([X=i] \cap [Y=k-i])$$

2) Loi géométrique

Nous considérons une épreuve à deux issues dont la probabilité du succès est p et celle de l'échec est notée $q=1-p$. Nous répétons cette épreuve de manière indépendante jusqu'au premier succès. Nous définissons la variable aléatoire X qui donne le rang du premier succès. Cette loi est notée $\mathcal{G}(p)$.

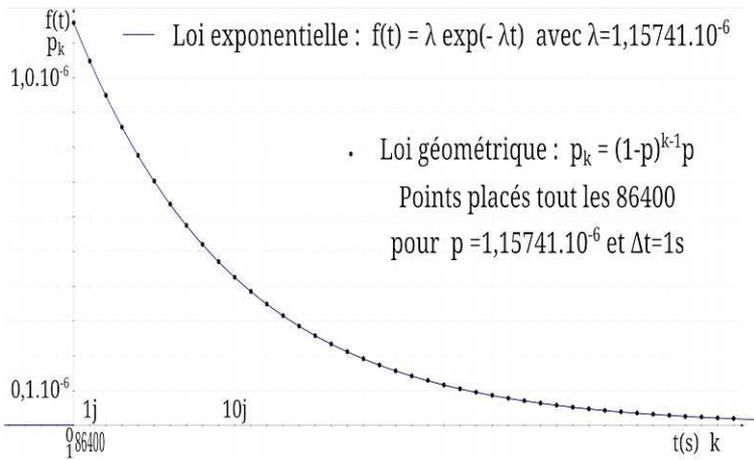
$$P(X=k)=q^{k-1} p \quad , \quad \sum_{k=1}^{\infty} P(X=k)=1$$

$$E(X)=1/p \quad \text{et} \quad V(X)=q/p^2$$

• Exemple : Nous avons une urne comportant 2 boules blanches et une boule noire. Les boules sont indiscernables au toucher. Quelle est la probabilité de tirer la première boule noire au troisième tirage avec remise ?

Réponse : $P(X=3)=(2/3)^2 1/3=4/27$.

- Application : nous avons ici un modèle discret de durée de vie. Prenons par exemple une cellule qui chaque jour a un chance sur dix de mourir. Si elle n'est pas morte le lendemain, sa probabilité de mourir le jour suivant n'a pas changée : il n'y a pas de vieillissement. Chaque jour elle peut échapper à la mort et elle peut vivre éternellement. De manière équivalente, plutôt que de prendre une décision chaque jour, nous pouvons décider de sa mort toutes les heures avec une probabilité q' de survivre telle que $q'^{24}=q$ soit $q' \simeq 0,996$. Ou même un choix toutes les secondes, nous avons alors des lois géométriques qui modélisent une même réalité avec $p \rightarrow 0$ et $q \rightarrow 1$.



Nous passons d'une loi discrète à une loi continue, le temps écoulé depuis le début de l'expérience est $t=k \Delta t$.

Regardons la fonction de répartition :

$$P(X \leq k) = \sum q^{k-1} p = \sum e^{\ln(1-p) \cdot (k-1)} p \simeq \sum p e^{-pk}$$

Nous avons utilisé un développement limité, le fait que $t \gg \Delta t$ et $k \gg 1$.

Dans le cas limite :

$P(X \leq k) \rightarrow \int f(t) dt$, $\sum p e^{-\frac{p}{\Delta t} t} \rightarrow \int \lambda e^{-\lambda t} dt$ ainsi la loi exponentielle est la loi continue équivalente à la loi géométrique en prenant $\lambda = p/\Delta t$, ici pour notre cellule λ vaut environ $1,16 \cdot 10^{-6}$ par seconde.

La somme de deux variables aléatoires géométriques indépendantes $G(p)$ est une loi binomiale négative $\mathcal{BN}(2, p)$, rang pour l'obtention de deux succès.

Pour une loi binomiale le nombre d'épreuves est fixé et nous regardons le nombre de succès, dans le cas de la loi binomiale négative c'est l'inverse nous nous intéressons au nombre d'épreuves nécessaires pour réaliser un nombre de succès fixé à l'avance. $\mathcal{BN}(r, p)$ est la loi de probabilité du rang du r^{em} succès¹⁹. Ainsi $G(p) = \mathcal{BN}(1, p)$.

3) Loi de Poisson

Soient des événements qui se produisent avec une fréquence connue indépendante du temps écoulé depuis les événements précédents. Si λ est le nombre de fois que l'événement se produit en moyenne sur un intervalle donné alors la probabilité que l'événement se produise k fois sur

¹⁹ Il existe aussi une autre définition : nombre d'échecs précédents le r^{em} succès. Les relations restent vraies en redéfinissant de même la loi géométrique.

cet intervalle de temps est :

$$P(X=k) = \frac{\lambda^k}{k!} e^{-\lambda}, \quad \sum_{k=0}^{\infty} P(X=k) = 1$$

$$E(X) = \lambda \quad \text{et} \quad V(X) = \lambda$$

- Exemple : Lors des Perséides d'août 1911 le taux horaire d'étoiles filantes était de 50, qu'elle était alors la probabilité de voir exactement 7 météores en 12 minutes ?

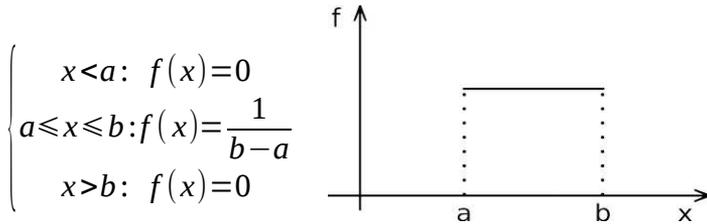
Réponse :

$$\lambda = \frac{12}{60} 50 = 10 \quad \text{et} \quad P(X=7) = \frac{10^7}{7!} e^{-10} \simeq 9\%$$

La somme de deux variables aléatoires de Poisson indépendantes $\mathcal{P}_X(\lambda_1)$ et $\mathcal{P}_Y(\lambda_2)$ est une loi de Poisson $\mathcal{P}_Z(\lambda_1 + \lambda_2)$.

B. Lois continues

1) Loi uniforme



$$E(X) = \frac{a+b}{2} \quad \text{et} \quad V(X) = \frac{(b-a)^2}{12}$$

Pour déterminer la loi de Z somme de X et Y variables aléatoires indépendantes continues nous utilisons un produit de convolution :

$$f_Z(x) = \int_{-\infty}^{+\infty} f_X(y) f_Y(x-y) dy$$

Considérons la somme de deux lois uniformes continues indépendantes $\mathcal{U}(a,b)$. Ici l'intégrand est non nul si :

$$a < y < b \quad \text{et} \quad a < x-y < b \quad \text{soit} \quad x-b < y < x-a$$

Si $2a < x < a+b$:

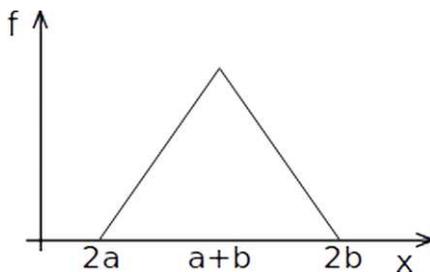
Si $a+b < x < 2b$:

Si $2a < x < a+b$

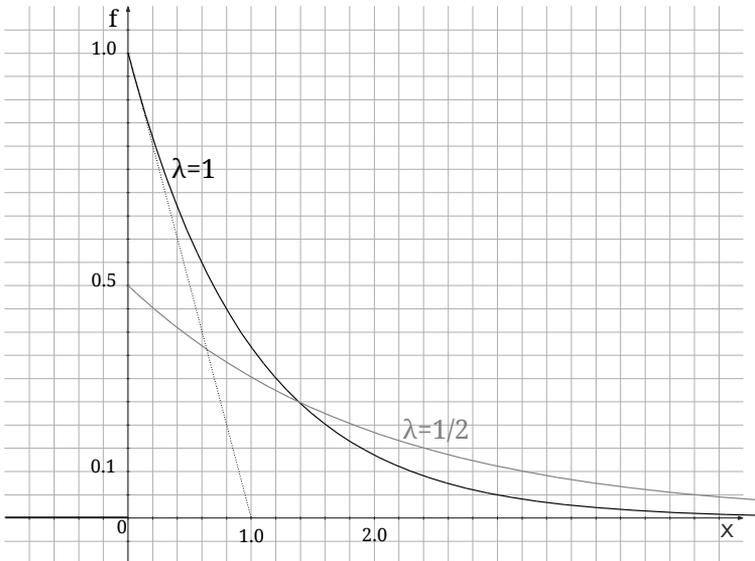
alors $f_z(x) = \int_a^{x-a} \frac{1}{(b-a)^2} dy = \frac{x-2a}{(b-a)^2}$

$$\left\{ \begin{array}{l} x < 2a : f_z(x) = 0 \\ 2a \leq x < a+b : f_z(x) = \frac{x-2a}{(b-a)^2} \\ a+b \leq x \leq 2b : f_z(x) = \frac{2b-x}{(b-a)^2} \\ x > 2b : f_z(x) = 0 \end{array} \right.$$

Nous obtenons une loi triangulaire :



2) Loi exponentielle



$$\begin{cases} t \geq 0: f(t) = \lambda e^{-\lambda t} \\ t < 0: f(t) = 0 \end{cases} \quad E(T) = \frac{1}{\lambda} \quad \text{et} \quad V(T) = \frac{1}{\lambda^2}$$

Loi de durée de vie sans vieillissement, en effet, on montre pour cette loi que : $P(T > b) = P_{T > a}(T > a + b)$.

Bien différencier la vie moyenne $E(T)$ avec la demi-vie $t_{1/2}$ telle que $P(T > t_{1/2}) = 0,5$.

La loi de la somme de deux lois exponentielles indépendantes n'est pas une loi exponentielle.

3) Loi normale

La loi normale, encore nommée gaussienne, a précédemment été décrite page 20.

La somme de deux variables aléatoires normales indépendantes $\mathcal{N}_X(\mu_1, \sigma_1^2)$ et $\mathcal{N}_Y(\mu_2, \sigma_2^2)$ est une loi normale $\mathcal{N}_Z(\mu_1 + \mu_2, \sigma_1^2 + \sigma_2^2)$.

4) Loi de Student

La loi de Student dépend du nombre de degrés de liberté k et utilise la fonction Gamma, fonction particulière décrite dans les outils mathématiques.

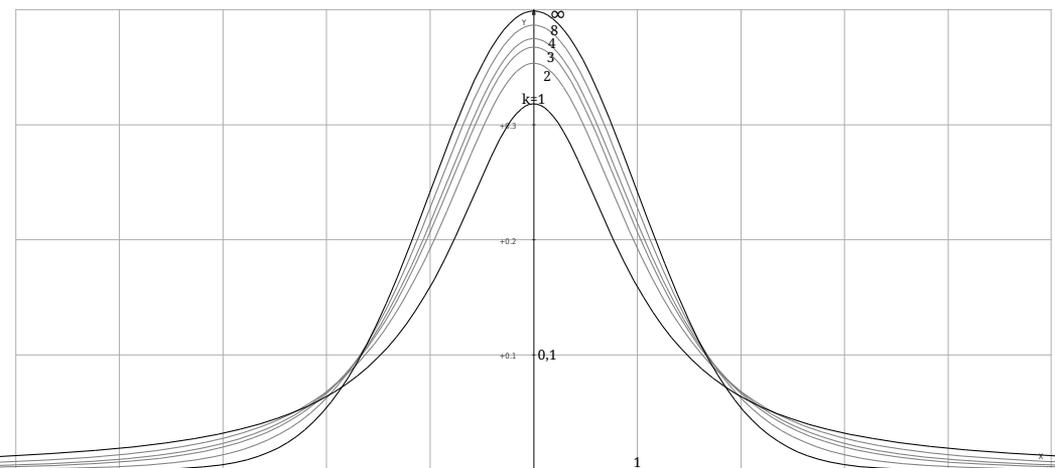
$$\text{Pour } k \geq 1, \quad f_k(x) = \frac{1}{\sqrt{k\pi}} \frac{\Gamma\left(\frac{k+1}{2}\right)}{\Gamma\left(\frac{k}{2}\right)} \frac{1}{\left(1 + \frac{x^2}{k}\right)^{\frac{k+1}{2}}}$$

La loi de Student présentée tend vers la loi normale centrée réduite $\mathcal{N}(0,1)$ quand k tend vers l'infini :

$$\lim_{k \rightarrow +\infty} f_k(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}$$

$$\text{Variance : } V_k = \frac{k}{k-2}, \text{ pour } k \geq 3.$$

$$\text{Coefficient d'aplatissement : } \beta_k = 3 \frac{k-2}{k-4}, \text{ pour } k \geq 5.$$



En exercices les expressions des premières Students sont exprimées, nous montrons que $\int_{-\infty}^{+\infty} f_k(x) dx = 1$, nous calculons l'expression de la variance et finalement nous montrons sur des exemples que la somme de deux Students indépendantes n'est pas une Student.

5) Loi du Khi-Deux

Soit k lois normales $\mathcal{N}(0,1)$ T_1, T_2, \dots et T_k indépendantes. La somme $X_k = T_1^2 + T_2^2 + \dots + T_k^2$ obéit à la loi du χ^2 . La loi du Khi^2 dépend du nombre de degrés de liberté k

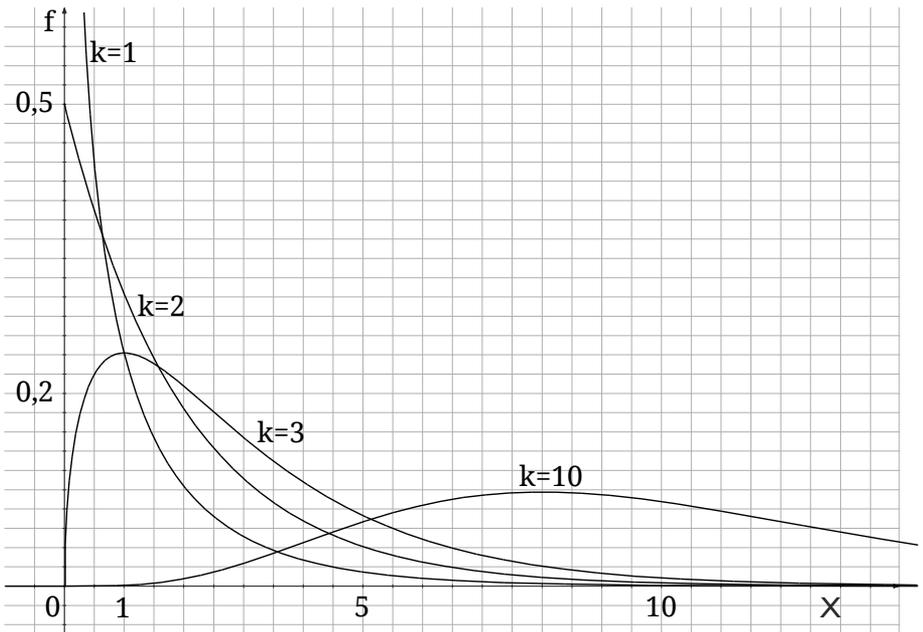
$$E(X_k) = \sum_{i=1}^k E(T_i^2) = k(V(T) + E(T)^2) = k$$

Pour $k \geq 1$ et $x \geq 0$, $f_k(x) = \frac{1}{2^{k/2} \Gamma(k/2)} x^{k/2-1} e^{-x/2}$

Espérance : $E_k = k$ Variance : $V_k = 2k$

Coefficient d'asymétrie : $\beta_{1,k} = \sqrt{8/k}$

Coefficient d'aplatissement : $\beta_{2,k} = 3 + 12/k$



La loi du Khi-deux tend vers la loi normale d'espérance k et de variance $2k$ quand k est grand.

C. Fonctions de variables à densité

Nous avons une variable aléatoire Y définie comme une fonction d'une variable aléatoire X continue : $Y=\varphi(X)$.

Nous connaissons la loi de X et nous voulons déterminer celle de Y . Nous raisonnons sur les fonctions de répartition F et nous dérivons ensuite pour avoir la densité f .

Fonction de répartition de X :

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

Nous considérons le cas où $\varphi(x)$ est une fonction strictement monotone. La loi de Y est dans ce cas continue. Fonction de répartition de Y :

$$F_Y(y) = P(Y \leq y) \quad \text{puis} \quad f_Y(y) = \frac{d F_Y(y)}{dy}$$

Nous cherchons à exprimer $F_Y(y)$ en fonction de F_X .

1) Cas où $\varphi(x)=\ln(x)$: $x>0$ et $Y=\ln X$

$$F_Y(y) = P(Y \leq y) = P(\ln X \leq y) = P(e^{\ln X} \leq e^y) = P(X \leq e^y)$$

Nous avons appliqué la fonction réciproque $\varphi^{-1}(x)=e^x$.

La fonction exponentielle est strictement croissante, le sens de l'inégalité a donc été conservé.

Ainsi : $F_Y(y) = F_X(e^y)$ et $f_Y(y) = e^y f_X(e^y)$

• Exemple : la loi de X est la loi uniforme $\mathcal{U}(1;2)$. Quelle est la loi de Y ?

$$\text{Nous avons : } f_X(x) = \begin{cases} 0 & x \leq 1 \\ 1 & 1 < x < 2, \text{ ainsi si } e^y \leq 1, \\ 0 & x \geq 2 \end{cases}$$

$y \leq 0$ et $f_Y(y) = 0$. Nous continuons ainsi pour les deux autres cas et nous avons la loi de Y :

$$f_Y(y) = \begin{cases} 0 & y \leq 0 \\ e^y & 0 < y < \ln 2 \\ 0 & y \geq \ln 2 \end{cases}$$

2) Cas où $\varphi(x) = ax + b$: $a \neq 0$ et $Y = aX + b$

Si $a > 0$:

$$F_Y(y) = P(Y \leq y) = P(aX + b \leq y) = P\left(X \leq \frac{y-b}{a}\right)$$

La fonction affine est strictement croissante pour a strictement positif, le sens de l'inégalité est conservé.

$$\text{Ainsi : } F_Y(y) = F_X\left(\frac{y-b}{a}\right) \text{ et } f_Y(y) = \frac{1}{a} f_X\left(\frac{y-b}{a}\right)$$

Si $a < 0$:

$$F_Y(y) = P\left(X \geq \frac{y-b}{a}\right) = 1 - P\left(X < \frac{y-b}{a}\right) = 1 - F_X\left(\frac{y-b}{a}\right)$$

$$\text{et } f_Y(y) = -\frac{1}{a} f_X\left(\frac{y-b}{a}\right) \text{ d'où } f_Y(y) = \frac{1}{|a|} f_X\left(\frac{y-b}{a}\right)$$

- Exemple 1 : la loi de X est la loi uniforme $\mathcal{U}(0;1)$. Quelle est la loi de Y ?

$$\text{Nous avons : } f_X(x) = \begin{cases} 0 & x \leq 0 \\ 1 & 0 < x < 1, \text{ ainsi si } \frac{y-b}{a} \leq 0 \text{ et} \\ 0 & x \geq 1 \end{cases}$$

$a > 0$ alors $y \leq b$ et $f_Y(y) = 0$. Nous continuons ainsi pour les deux autres cas et nous avons la loi de Y :

$$f_Y(y) = \begin{cases} 0 & y \leq b & U(0,1) \rightarrow U(b, a+b) \\ \frac{1}{a} & b < y < a+b & \text{Si } \varphi(x) = (b-a)x + a \text{ et } a < b: \\ 0 & y \geq a+b & U(0,1) \rightarrow U(a, b) \end{cases}$$

- Exemple 2 : la loi de X est la loi normale $\mathcal{N}(0;1)$. Nous pouvons retrouver la loi d'une $\mathcal{N}(\mu;\sigma)$ avec $Y = \sigma X + \mu$. De manière générale par application d'une fonction affine nous obtenons une loi de même type.

3) Cas où $\varphi(x) = x^2$: $Y = X^2$ et $y > 0$

$$F_Y(y) = P(X^2 \leq y) = P(-\sqrt{y} \leq X \leq \sqrt{y}) = F_X(\sqrt{y}) - F_X(-\sqrt{y})$$

$$\text{et } f_Y(y) = \frac{1}{2\sqrt{y}} (f_X(\sqrt{y}) + f_X(-\sqrt{y})) \text{ si } y > 0 \text{ et zéro sinon.}$$

4) Cas où $\varphi(x)=e^x$: $Y=e^X$ et $y>0$

$$F_Y(y)=P(e^X \leq y)=P(X \leq \ln y)=F_X(\ln y)$$

$$\text{et } f_Y(y)=\frac{1}{y} f_X(\ln y) \text{ si } y>0 \text{ et zéro sinon.}$$

D. Simulation numérique

Nous simulons à l'aide d'ordinateurs des lois de probabilités continues et discrètes. Pour cela nous utilisons des lois uniformes créées par des algorithmes de génération de nombres aléatoires :

Lois uniformes continues $U(0,1)$: Ran# sur calculatrice, ALEA() sur LibreOffice, etc.

Lois uniformes discrètes $U(i,j)$: par exemple, rand(i,j) en langage PHP. rand(0,999)/1000 simule une loi uniforme continue discrétisée au millième.

- Méthode d'inversion : détermination de la loi de X en fonction de celle de U par inversion de la fonction de répartition. Avec F strictement croissante : $X=F^{-1}(U)$.

Cas d'une loi exponentielle : $f_X(x)=\lambda e^{-\lambda x}$ pour $x>0$ et zéro sinon, d'où $F_X(x)=\int_{-\infty}^x f_X(x) dx=1-e^{-\lambda x}=y$ pour

$x > 0$ et zéro sinon. Ainsi $x = -\ln \frac{(1-y)}{\lambda}$ et finalement nous simplifions, sachant que $1-U$ et U ont la même loi.

Simulation d'une loi exponentielle : $X = -\frac{\ln U}{\lambda}$

- Simulation de deux lois normales $\mathcal{N}(0;1)$ indépendantes X_1 et X_2 à partir de U_1 et U_2 indépendantes :

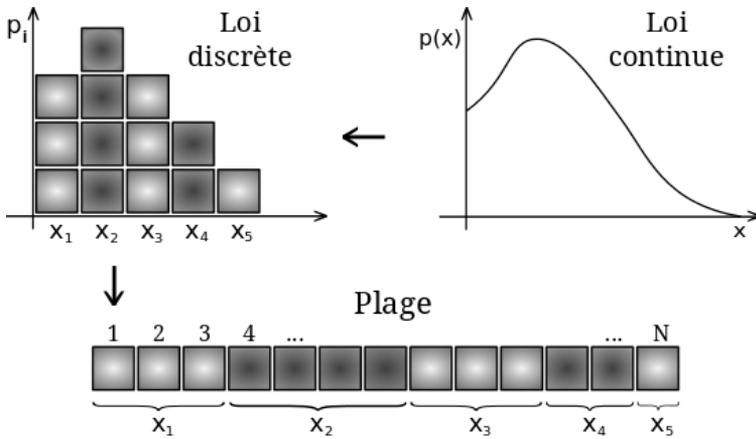
$$X_1 = \sqrt{-2 \ln U_1} \cos(2\pi U_2)$$

$$X_2 = \sqrt{-2 \ln U_1} \sin(2\pi U_2)$$

C'est la méthode de Box-Müller. Pour une gaussienne il n'y a pas de formulation directe de la fonction de répartition F . Ainsi F^{-1} n'a pas d'expression analytique simple.

- Méthode pour une loi X quelconque :

Si nous avons une loi continue nous discrétisons x par classes. Nous avons alors un histogramme et chaque barre de hauteur p_i est aussi discrétisée en unités. Une réalisation de la loi revient alors à tirer au hasard et de manière équiprobable une unité de l'histogramme. Nous obtenons en mettant les barres bout à bout une plage complète des unités. Nous utilisons une loi uniforme discrète $U(1,N)$ où N correspond au nombre total d'unités de l'histogramme. La valeur générée est comparée à sa position sur la plage et nous en déduisons une valeur pour la réalisation x_i .



- Il existe de nombreuses autres méthodes qui utilisent les différentes propriétés des lois de probabilité. Par exemple, en simulant des lois de Bernoulli²⁰, nous obtenons, par somme, une loi binomiale qui elle-même permet de simuler une loi normale.

20 Loi à deux issues : $p=P(S)$ et $q=1-p=P(E)$.

E. Exercices

Exercice 1 : Loi binomiale

corrigé en version complète

Vérifier qu'il s'agit d'une loi de probabilité. Déterminer l'espérance et la variance d'une loi binomiale en fonction de n et p .

Exercice 2 : Somme de binomiales

corrigé en version complète

Montrer que la somme de loi binomiales indépendantes de même paramètres p est elle-même une loi binomiale dont on déterminera les paramètres.

Exercice 3 : Loi géométrique

corrigé en version complète

Montrer qu'il s'agit bien d'une loi de probabilité. Déterminer l'espérance et la variance d'une loi géométrique en fonction de p .
Montrer que la somme de deux variables aléatoires géométriques indépendantes de paramètre p est une loi binomiale négative $\mathcal{N}(\mathcal{B}(2,p))$.

Exercice 4 : Premiers succès

corrigé en version complète

1- Soit une pièce équilibrée. Qu'elle est la probabilité que le premier Pile apparaisse au cinquième lancer ? Sachant que le premier Pile n'ait toujours pas apparu au troisième lancer, qu'elle est la probabilité qu'il apparaissent pour la première fois au huitième lancer ?

2- Soit un dé équilibré. En moyenne après combien de lancers apparaît le premier six ? Qu'elle est la probabilité que le premier six apparaisse au cours des six premiers lancers ?

Exercice 5 : Loi de Poisson

corrigé en version complète

Montrer qu'il s'agit bien d'une loi de probabilité.
Déterminer l'espérance et la variance d'une loi Poisson en fonction du paramètre λ .
Montrer la loi de la somme.

Exercice 6 : Loi uniforme

corrigé en version complète

Retrouver l'expression de la variance pour une loi uniforme continue $\mathcal{U}(a,b)$.

Exercice 7 : Loi exponentielle

corrigé en version complète

Montrer qu'il s'agit bien d'une loi de probabilité.
Déterminer l'espérance et la variance d'une loi exponentielle en fonction du paramètre λ .
Trouver la loi de la somme.

Exercice 8 : Somme de gaussiennes

corrigé en version complète

- 1- Trouver la loi de la somme de deux $\mathcal{N}(0,1)$.
- 2- Trouver la loi de la somme dans le cas général.

Exercice 9 : Premières Students

corrigé en version complète

- 1- Donner les expressions des premières fonctions de Student.
- 2- Donner le degré du polynôme au dénominateur, la variance, le coefficient d'aplatissement et la valeur en zéro.
- 3- Donner l'expression de la Student pour $k=9$, centrée et avec une variance de $14/5$.

Exercice 10 : Loi de Student

corrigé en version complète

Quelque soit k , montrer que la loi de Student correspond bien a une loi de probabilité.

On pourra poser l'intégrale $I_k = \int_{-\infty}^{+\infty} \left(1 + \frac{x^2}{k}\right)^{-\frac{k+1}{2}} dx$ et

effectuer le changement de variable $u = \frac{x}{\sqrt{k}}$.

Exercice 11 : Variance d'une Student

corrigé en version complète

Déterminer la variance d'une la loi de Student.

Exercice 12 : Somme de Students

corrigé en version complète

Sur un exemple montrer que la somme de deux variables aléatoires de Student indépendantes n'est pas une variable aléatoire de Student.

Exercice 13 : Lois du Khi-Deux

corrigé en version complète

Donner les expressions des premières fonctions du Khi-Deux pour $k=1$, $k=2$, $k=3$ puis $k=10$.

Exercice 14 : Loi du produit

corrigé en version complète

Soient deux lois X et Y indépendantes.

1. Proposer une méthode générale permettant de déterminer la loi de probabilité de $Z=XY$.
2. Nous considérons le cas où X et Y sont des lois uniformes continues $\mathcal{U}(1;2)$ indépendantes.
 - a. Déterminer l'expression analytique de la loi de Z .
 - b. Retrouver l'allure de la loi de probabilité de Z avec une simulation du produit sur tableur pour $n=10\ 000$.

Exercice 15 : Loi de la somme de lois exponentielles

corrigé en version complète

Soient n lois exponentielles X_i de même paramètre λ et indépendantes. Nous appelons S_n la loi de la somme : $S_n = X_1 + X_2 + \dots + X_n$. Nous avons aussi la loi : $M_n = S_n/n$.

1. Déterminer la loi de probabilité de S_2 .
2. Déterminer la loi de probabilité de S_n .
3. Déterminer la loi de probabilité de M_n .

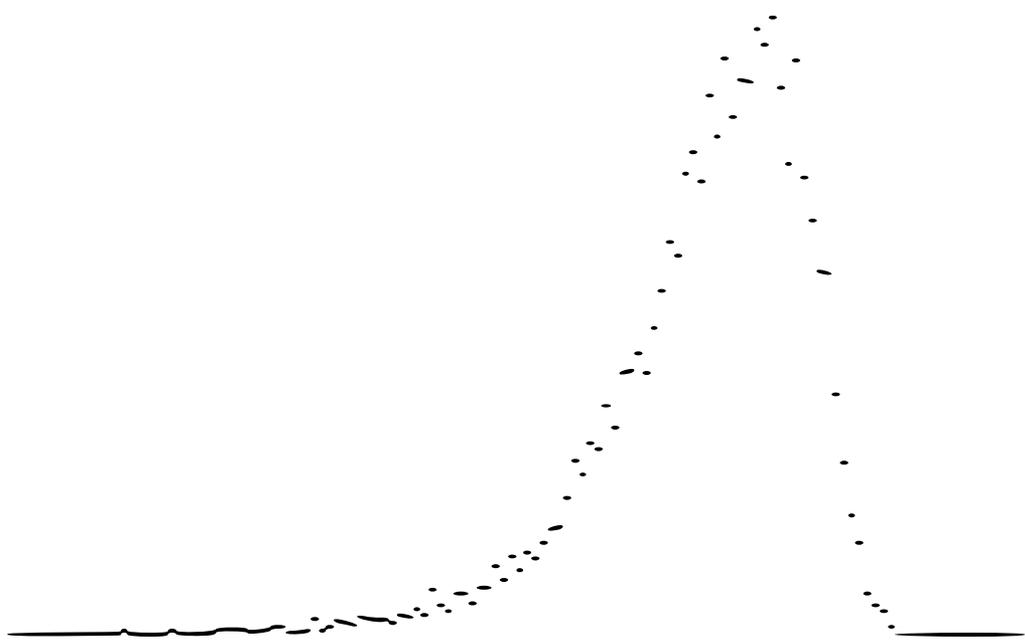
Exercice 16 : Loi de l'inverse

corrigé en version complète

Soit X une loi de probabilité définie strictement positive.

1. Déterminer la loi de $Y=1/X$.
2. a. Quelle la loi de T_n inverse de M_n défini à l'exercice précédent ?
b. Quelle la loi de l'inverse de la loi de Cauchy :

$$\text{loi de } X : f_x = \frac{1}{\pi} \frac{1}{1+x^2} \text{ pour tout } x.$$



IV. ESTIMATEURS

Nous avons une loi de probabilité X qui dépend d'un certain nombre de paramètres.

À l'aide d'un échantillon (X_1, X_2, \dots, X_n) de variables aléatoires mutuellement indépendantes et qui suivent la loi de X , nous voulons disposer d'une méthode pour estimer au mieux les différents paramètres qui définissent notre loi (statistique inférentielle). Notre échantillon est de taille n . Nos paramètres sont de manière générique désignés par la lettre θ et nous nommons T_n notre estimateur de θ .

Nous effectuons dans un premier temps une estimation ponctuelle, valeur la plus probable du paramètre, dans un second temps nous déterminerons une estimation par intervalle.

A. Qualité d'un estimateur

1) Biais

Nous appelons biais de T_n l'espérance $E(T_n - \theta)$ d'où, par linéarité de l'espérance :

$$b_{T_n}(\theta) = E(T_n) - \theta$$

L'idéal est d'avoir un estimateur sans biais ou sinon un biais qui tend vers zéro quand la taille de l'échantillon tend vers l'infini.

2) Risque

Nous appelons risque quadratique de T_n l'espérance $E[(T_n - \theta)^2]$ et on montre que :

$$r_{T_n}(\theta) = V(T_n) + b^2$$

En effet $E[(T_n - \theta)^2] = E(T_n^2) - 2\theta E(T_n) + \theta^2$ (par linéarité de l'espérance), et finalement après simplifications de $r = V(T_n) + E(T_n)^2 - 2\theta(b + \theta) + \theta^2$ nous retrouvons l'expression du risque. À moins d'un estimateur non biaisé le risque n'est pas simplement égal à la variance de l'estimateur. On a intérêt à choisir un estimateur dont le risque tend vers zéro avec la taille de l'échantillon et plus la convergence est rapide, meilleur est l'estimateur.

• Exemple 1 : Nous appelons \bar{X}_n la moyenne empirique associée à l'échantillon (X_1, X_2, \dots, X_n) définie par :

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

Montrez que \bar{X}_n est un estimateur de l'espérance $\theta = m = E(X)$. Déterminer la qualité de cette estimateur.

$E(T_n) = 1/n \sum E(X_i)$ par linéarité de l'espérance, d'où $E(T_n) = 1/n \cdot n m = m$ et $b = 0$.

$V(T_n) = 1/n^2 \sum V(X_i)$ (propriété de la variance et indépendance des variables), d'où $r = 1/n^2 \cdot n \sigma^2$ et $r = \sigma^2/n$.

La moyenne empirique est un bon estimateur de l'espérance d'une variable aléatoire. L'estimateur est de

biais nul et le risque tend vers zéro quand n tend vers l'infini.

• Exemple 2 : Nous cherchons un estimateur de la variance d'une variable aléatoire, nous proposons les deux estimateurs suivants :

$$S_n^2 = \frac{\sum_{i=1}^n (X_i - \bar{X}_n)^2}{n} \quad \text{et} \quad R_n^2 = \frac{\sum_{i=1}^n (X_i - \bar{X}_n)^2}{n-1}$$

Lequel, selon vous, est le meilleur estimateur de σ^2 ?

$$S_n^2 = 1/n \sum (X_i^2 - 2X_i\bar{X}_n + \bar{X}_n^2) \quad \text{et}$$

$$S_n^2 = 1/n \sum X_i^2 - 2T_n \sum X_i/n + T_n^2 = 1/n \sum X_i^2 - T_n^2$$

$$\text{D'où } E(S_n^2) = E(X^2) - E(T_n^2) = \sigma^2 + m^2 - V(T_n) - E(T_n)^2$$

$$\text{Et } E(S_n^2) = \sigma^2 - \frac{\sigma^2}{n} \quad \text{soit } E(S_n^2) = \frac{n-1}{n} \sigma^2 \quad \text{et } b_{S_n^2} = -\frac{\sigma^2}{n}.$$

Par un calcul tout à fait analogue nous trouvons $b_{R_n^2} = 0$.

Les deux quantités sont des estimateurs de la variance car les biais sont, soit nul, soit asymptotiquement non biaisé.

R_n^2 est le meilleur estimateur car il n'a pas de biais.

B. Construction d'estimateurs

Nous voulons disposer d'une méthode générale permettant de trouver les estimateurs adéquats pour les paramètres d'une loi.

1) Méthode des moments

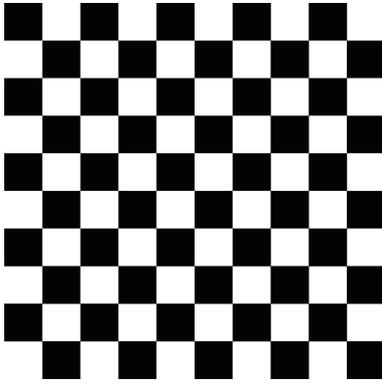
Nous identifions les moments de la population avec ceux de l'échantillon. Nous considérons autant de moments que nous avons de paramètres en commençant par le moment d'ordre 1. Comme nous le verrons sur des exemples cette méthode nous fournit des estimateurs mais ne nous garantit pas que ce sont les meilleurs en terme de biais et de risque quadratique.

$$\text{Moments de la loi : } m_k = E(X^k)$$

$$\text{Moments de l'échantillon : } \overline{X}_n^k = \frac{1}{n} \sum_{i=1}^n X_i^k$$

Nous pouvons aussi considérer les moments centrés ou même centrés réduits, l'approche reste la même. Le moment du n-échantillon d'ordre 1 est la moyenne empirique et correspond bien à un excellent estimateur comme montré précédemment. Par contre le moment centré d'ordre 2 a un biais, comme nous l'avons montré avant il faudrait diviser par $n-1$ au lieu de n pour avoir un biais nul.

- Exemple 1 : Nous disposons d'un damier de 100 cases et de 200 graines. Chaque seconde nous plaçons aléatoirement une graine dans une des cases. À la fin de l'expérience nous comptons le nombre de graines dans chaque case et nous dénombrons le nombre de cases qui ne contiennent aucune graine, puis une graine, puis deux et ainsi de suite. Soit la loi



X du nombre de graines par case. Nous obtenons la distribution suivante :

k	0	1	2	3	4	5	6	7	8
n	12	28	34	10	8	7	0	1	0

Nous supposons que cette distribution suit une loi de Poisson, en déduire la valeur du paramètre λ de la loi.

Le paramètre de cette loi est égal à l'espérance donc d'après le théorème de moments, λ est estimé par la moyenne

$$\text{empirique : } \lambda = \frac{1}{n} \sum_{i=1}^n x_i = \frac{1}{n} \sum_{k=1}^p n_k x_k \text{ soit } \lambda = 2 \text{ }^{21}.$$

Résultat cohérent avec notre modèle, en effet nommons n le nombre de cases et N le nombre de graines. Nous avons une distribution aléatoire uniforme. Nous pourrions par exemple utiliser deux dés équilibrés dix faces de couleurs différentes pour la position horizontale et verticale. La

²¹ En utilisant le moment d'ordre 2 nous trouverions $\lambda = 2,06$.

fréquence de l'événement par seconde pour chaque case vaut $1/n$. Pour une loi de Poisson à tout instant l'événement peut se produire, ici le temps est discrétisé, néanmoins l'approximation d'un temps continu est correct car une seconde est une petite durée au regard de celle de 200 secondes de l'expérience. Nous pouvons donc prendre $\lambda = N/n$. Nous tendons vers une loi de Poisson quand le nombre des cases et de graines tendent vers l'infini.

- Exemple 2 : Nous supposons que la durée de vie d'un verre de table suit une loi exponentielle. Nous observons les différentes durées de vie en mois suivantes : 7, 5, 37, 35, 17, 9, 6, 13, 4 et 8 mois. Déterminez le paramètre λ de la loi.

Pour un loi exponentielle $E(T) = \frac{1}{\lambda}$ d'où en utilisant le moment d'ordre 1 : $\lambda = \frac{n}{\sum_{i=1}^n t_i}$ et $\lambda = \frac{10}{141} \approx 0,071$.

Nous avons ainsi une estimation ponctuelle du paramètre λ et l'espérance de vie vaut environ 14 mois.

Nous démontrerons en exercice que cet estimateur T_n possède un biais : $E(T_n) = \frac{n}{n-1} \lambda$ et $b_{T_n} = \frac{\lambda}{n-1}$.

Ce biais est asymptotiquement nul et son risque tend vers zéro : $r_{T_n} = \lambda^2 \frac{(n+2)}{(n-1)(n-2)}$.

Nous construisons à partir de cet estimateur un nouvel estimateur W_n sans biais : $W_n = \frac{n-1}{n} T_n = \frac{n-1}{\sum X_i}$.

On montre : $E(W_n) = \lambda$ et $r_{W_n} = V(W_n) = \lambda^2 / (n-2)$.

Nous avons un meilleur estimateur : biais nul et risque plus faible. Une nouvelle estimation du paramètre donne $\lambda \simeq 0,064$ et l'espérance de vie vaut environ 16 mois.

• Exemple 3 : Nous considérons que la masse des pommes produites par un arbre suivent une loi normale. Nous prélevons au hasard et mesurons les masses suivantes en grammes :

158	131	146	158	125	153	166	121
127	123	195	149	124	153	123	129

Déterminer la moyenne et l'écart-type de la loi.

$$\mu = \frac{1}{n} \sum_{i=1}^n m_i \simeq 142,5, \quad \sigma^2 + \mu^2 = \frac{1}{n} \sum_{i=1}^n m_i^2 \quad \text{et} \quad \sigma \simeq 20,2.$$

• Exemple 4 : Les planches produites par une scierie ont une longueur qui suivent une loi uniforme $\mathcal{U}(a, b)$.

Nous mesurons les longueurs en millimètres de 8 planches tirées au hasard : 2017, 1987, 2018, 2014, 2003, 1985, 2013 et 1981. Estimez a et b.

$$\text{Nous posons} \quad E(X) = \frac{a+b}{2} = \overline{X}_n$$

$$\text{et} \quad E(X^2) = \frac{(b-a)^2}{12} + \frac{(a+b)^2}{4} = \frac{a^2 + ab + b^2}{3} = \overline{X}_n^2$$

$$\text{d'où} \quad a+b = 2\overline{X}_n \quad \text{et} \quad ab = 4\overline{X}_n^2 - 3\overline{X}_n^2$$

Finalement : $a \simeq 1977$ et $b \simeq 2028$

2) Méthode du maximum de vraisemblance

Soit un variable aléatoire X discrète ou continue dont nous voulons estimer un paramètre θ à partir d'un ensemble d'observations $\{x_i\}$.

Nous définissons une fonction $f(x, \theta)$ telle que :

$$f(x, \theta) = \begin{cases} P_\theta(X=x) & \text{pour une variable discrète} \\ & \text{ou} \\ f_\theta(x) & \text{pour une variable continue} \end{cases}$$

La vraisemblance V est ainsi définie :

$$V(x_1, x_2, \dots, x_i, \dots, x_n, \theta) = f(x_1, \theta) \times f(x_2, \theta) \times \dots \times f(x_n, \theta)$$

notée aussi :

$$V(x_1, x_2, \dots, x_i, \dots, x_n, \theta) = \prod_{i=1}^n f(x_i, \theta)$$

La valeur du paramètre pour laquelle le produit des probabilités, ou densités de probabilités, prises aux différents points de l'échantillon est maximale, est considérée comme la valeur la plus vraisemblable.

Le principe est simple et la méthode est facile à mettre en œuvre :

$$\frac{\partial V(x_i, \theta)}{\partial \theta} = 0 \quad \text{et} \quad \frac{\partial^2 V(x_i, \theta)}{\partial \theta^2} < 0$$

La nullité de la dérivée permet d'obtenir notre estimateur : relation entre θ et les x_i .

La fonction logarithme étant strictement croissante nous pouvons aussi travailler avec $\ln V$ pour des calculs plus simples.

- Exemple 1 : Reprenons le cas d'une loi de Poisson et déterminons l'estimateur de λ par la méthode du maximum de vraisemblance.

$$P_\lambda(X=x) = \frac{\lambda^x}{x!} e^{-\lambda} \quad \text{et} \quad V = \prod_{i=1}^n \frac{\lambda^{x_i}}{x_i!} e^{-\lambda} = e^{-n\lambda} \prod_{i=1}^n \frac{\lambda^{x_i}}{x_i!}$$

$$\ln V = -n\lambda + \sum_{i=1}^n \ln \frac{\lambda^{x_i}}{x_i!} \quad \text{et} \quad \frac{\partial \ln V}{\partial \lambda} = -n + \sum_{i=1}^n \frac{x_i}{\lambda} = 0$$

$$\text{D'où} \quad \lambda = \frac{1}{n} \sum_{i=1}^n x_i$$

Nous trouvons le même estimateur que par la méthode précédente.

- Exemple 2 : Reprenons le cas d'une loi exponentielle et déterminons l'estimateur de λ par la méthode du maximum de vraisemblance.

$$f_\lambda(t) = \lambda e^{-\lambda t} \quad \text{et} \quad V = \prod_{i=1}^n \lambda e^{-\lambda t_i} = \lambda^n e^{-\lambda \sum_{i=1}^n t_i}$$

$$\ln V = n \ln \lambda - \lambda \sum_{i=1}^n t_i \quad \text{et} \quad \frac{\partial \ln V}{\partial \lambda} = \frac{n}{\lambda} - \sum_{i=1}^n t_i$$

$$\text{D'où} \quad \lambda = \frac{n}{\sum_{i=1}^n t_i}$$

Le même estimateur que par la méthode précédente.

- Exemple 3 : Reprenons le cas d'une loi normale et déterminons les estimateurs de μ et σ par la méthode du maximum de vraisemblance.

$$f(x) = \frac{1}{\sqrt{2\pi} \cdot \sigma} e^{-\frac{1}{2} \left(\frac{x-\mu}{\sigma} \right)^2} \quad \text{et} \quad V = \prod_{i=1}^n \frac{1}{\sqrt{2\pi} \sigma} e^{-\frac{1}{2} \left(\frac{x_i-\mu}{\sigma} \right)^2}$$

$$\text{d'où} \quad \ln V = -n \ln \sigma - \frac{n}{2} \ln(2\pi) - \frac{1}{2} \sum_{i=1}^n \left(\frac{x_i - \mu}{\sigma} \right)^2$$

$$\frac{\partial \ln V}{\partial \mu} = \sum_{i=1}^n \frac{x_i - \mu}{\sigma^2} = 0 \quad \text{et donc} \quad \mu = \frac{1}{n} \sum_{i=1}^n x_i$$

$$\frac{\partial \ln V}{\partial \sigma} = -\frac{n}{\sigma} + \frac{1}{\sigma^3} \sum_{i=1}^n (x_i - \mu)^2 = 0 \quad \text{et} \quad \sigma^2 = \frac{1}{n} \sum_{i=1}^n (x_i - \mu)^2$$

Nous obtenons les mêmes estimateurs que par le théorème des moments et nous avons à nouveau un estimateur de la variance biaisé.

- Exemple 4 : Reprenons le cas de la loi uniforme.

$$V(x_i, a, b) = \prod_{i=1}^n f(x_i, a, b) = \begin{cases} \frac{1}{(b-a)^n} & \text{si } \{x_i\} \subset [a, b] \\ 0 & \text{sinon} \end{cases}$$

Pour a fixé, plus b est petit plus la vraisemblance est grande donc $b = \max(\{x_i\})$.

Pour b fixé, plus a est grand plus la vraisemblance est grande donc $a = \min(\{x_i\})$.

Nous avons ici des estimateurs bien différents que par la méthode des moments, pour l'exemple des planches nous avons les estimations suivantes : $a = 1981$ et $b = 2018$.

Ces estimateurs sont biaisés, par exemple b est nécessairement inférieur à la valeur théorique, mais contrairement à la méthode des moments nous sommes assuré que les x_i appartiennent à $[a, b]$.

C. Estimation par intervalle

Nous avons maintenant des outils efficaces pour déterminer la valeur des paramètres d'une loi. Nous avons déterminé des valeurs ponctuelles et à présent nous voulons déterminer un intervalle de confiance.

Le théorème central limite permet à partir d'un échantillon de grande taille d'estimer la moyenne d'une loi de probabilité avec un intervalle de confiance.

Mais qu'en est-il si nous voulons déterminer d'autres paramètres différents de la moyenne ? Par exemple, quelles sont les incertitudes sur le paramètre λ d'une loi exponentielle, ou les bornes a et b pour une loi uniforme ?

Nous considérons un estimateur non biaisé et si l'estimateur est biaisé nous en créons un nouveau en retirant le biais.

Nous utilisons trois méthodes différentes. La méthode par intégration qui exige de déterminer la loi de probabilité complète de l'estimateur. Une seconde méthode par encadrement, plus simple, qui surestime notre incertitude, mais ne requiert que la connaissance de la variance de l'estimateur. Et finalement une troisième par simulation numérique.

L'inégalité de Bienaymé-Tchebychev fournit l'encadrement de la deuxième méthode. Elle est universelle car elle s'applique à toute loi X possédant une espérance et une variance. Nous devons cependant connaître la variance de notre estimateur et elle fournit un majorant assez grossier en utilisant la probabilité suivante :

$$\forall \epsilon > 0, \quad P(|X - E(X)| \geq \epsilon) \leq \frac{V(X)}{\epsilon^2}$$

• Exemple 1 : Considérons à nouveau le damier et la loi de Poisson de paramètre λ . Ce paramètre est estimé par la moyenne et nous pouvons donc, dans le cas des grands nombres, utiliser le théorème central limite :

$$\lambda = \lambda_m \pm t_\infty \sigma / \sqrt{n}$$

Estimons la variance :

k	0	1	2	3	4	5	6	7	8
$k - \lambda_m$	-2	-1	0	1	2	3	4	5	6
$(k - \lambda_m)^2$	4	1	0	1	4	9	16	25	36
n	12	28	34	10	8	7	0	1	0

$$\sigma_\lambda^2 = \frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})^2, \quad \sigma_\lambda \simeq 1,44 \text{ et } \lambda = 2,0 \pm 0,3 \text{ à } 95\%.$$

Ici la taille de l'échantillon n'est pas suffisante pour cette méthode mais nous avons illustré la méthode générale pour de grands échantillons. Nous avons probablement sous-estimé la largeur de l'intervalle.

• Deuxième exemple : Le temps d'attente d'un train suit une loi uniforme $\mathcal{U}(0, a)$. Nous observons les temps d'attente suivants 3, 12, 7, 17, 8, 14, 2, 5, 10, 14, 15 et 11 minutes. Que vaut a ?

Quel est le biais, la variance de l'estimateur T_n et l'incertitude sur a à 90% de confiance ?

a) Utilisons tout d'abord la méthode du maximum de vraisemblance qui nous fournit la loi du max pour connaître la loi de probabilité de T_n : $T_n = \max(\{X_i\})$.

a est donc ici estimé à 17 minutes.

Déterminons la loi de l'estimateur, loi du max :

$$P(T_n \leq x) = P([X_1 \leq x] \cap [X_2 \leq x] \cap \dots \cap [X_n \leq x])$$

or les variables sont indépendantes :

$$\text{d'où } P(T_n \leq x) = \prod_{i=1}^n P(X_i \leq x)$$

Avec les fonctions de répartition : $F_X(x) = P(X \leq x)$

$$F_{T_n}(x) = \prod_{i=1}^n F_{X_i}(x) = \left(\frac{x}{a}\right)^n \text{ pour } 0 \leq x \leq a$$

$$F_{T_n}(x) = 0 \text{ si } x < 0, F_{T_n}(x) = 1 \text{ si } x > a$$

$$\text{d'où la loi de probabilité de } T_n : f_{T_n}(x) = \frac{dF_{T_n}(x)}{dx} = n \frac{x^{n-1}}{a^n}$$

pour $0 \leq x \leq a$ et $f_{T_n}(x) = 0$ sinon.

$$\text{Espérance : } E(T_n) = \int x f_{T_n}(x) dx = \frac{n}{a^n} \int_0^a x^n dx = \frac{n}{n+1} a$$

$$\text{l'estimateur } T_n \text{ est biaisé : } b_{T_n} = \frac{n}{n+1} a - a = -\frac{a}{n+1}$$

a est sous-estimé et le biais est asymptotiquement nul.

$$\text{Variance : } E(X^2) = \int x^2 f_{T_n}(x) dx = \frac{n}{a^n} \int_0^a x^{n+1} dx = \frac{n}{n+2} a^2$$

$$V(T_n) = E(X^2) - E(X)^2 = \frac{n}{n+2} a^2 - \frac{n^2}{(n+1)^2} a^2 = \frac{na^2}{(n+2)(n+1)^2}$$

Il est préférable de prendre un estimateur non biaisé, pour cela nous retirons le biais et nous obtenons W_n :

$$W_n = \frac{n+1}{n} T_n \quad \text{avec } b_{W_n} = 0 \quad \text{et } V(W_n) = \frac{a^2}{n(n+2)}$$

Nouvelle estimation de a avec W_n : $a \simeq 18,4$ minutes.

Pour l'incertitude Δa nous allons comparer trois méthodes.

-> Déterminons un intervalle de confiance grâce à l'inégalité de Bienaymé-Tchebychev :

Nous appliquons l'inégalité à W_n :

$$P(|W_n - a| \geq \epsilon) \leq \frac{V(W_n)}{\epsilon^2}$$

$$\text{or } P(|W_n - a| \leq \epsilon) = P(-\epsilon \leq W_n - a \leq \epsilon)$$

$$\text{et } P(\epsilon \geq a - W_n \geq -\epsilon) = P(U_n + \epsilon \geq a \geq W_n - \epsilon)$$

Puis en posant $\epsilon = \sqrt{\frac{V(W_n)}{\alpha}}$ nous avons :

$$1 - P(W_n - \epsilon \leq a \leq W_n + \epsilon) \leq \alpha$$

Et finalement :

$$P\left(W_n - \sqrt{\frac{V(W_n)}{\alpha}} \leq a \leq W_n + \sqrt{\frac{V(W_n)}{\alpha}}\right) \geq 1 - \alpha$$

Pour un confiance de 90%, $\alpha=0,1$ et pour notre échan-

$$\text{tillon } V(W_{12}) = \frac{18,4^2}{12 \times 14} \simeq 2,0 \text{ et } \Delta a = \sqrt{\frac{V(W_n)}{\alpha}} \simeq 4,5 .$$

D'où $13,9 \leq a \leq 22,9$ et $a \simeq 18,4 \pm 4,5$ minutes.

-> Déterminons un intervalle de confiance avec un calcul d'intégrales sur la densité de probabilité de l'estimateur.

Nous déterminons la densité de probabilité de W_n :

$$P(W_n \leq x) = P\left(\frac{n}{n+1} W_n \leq \frac{n}{n+1} x\right) = P\left(T_n \leq \frac{n}{n+1} x\right)$$

$$F_{W_n}(x) = F_{T_n}\left(\frac{n}{n+1} x\right) = \left(\frac{n}{n+1} \frac{x}{a}\right)^n \text{ pour } 0 \leq x \leq \frac{n+1}{n} a .$$

$$f_{W_n}(x) = \left(\frac{n}{a(n+1)}\right)^n n x^{n-1} \text{ pour } 0 \leq x \leq \frac{n+1}{n} a$$

et $f_{W_n}(x) = 0$ sinon.

Nous avons une loi de probabilité dissymétrique dont le maximum correspond à la borne supérieure. Pour une confiance de 90% nous retirons les 10% de la queue de distribution gauche de l'intervalle de confiance. Ainsi nous définissons a_{max} et a_{min} tels que :

$$\text{Borne supérieure : } a_{max} = \frac{n+1}{n} a$$

soit $a_{max} = \left(\frac{13}{12}\right)^2 \times 17 \simeq 19,95$ et $a_{min} \simeq 20,0$ minutes.

Borne inférieure :
$$\int_{a_{min}}^{a_{max}} f_{W_n}(x) dx = 1 - \alpha$$

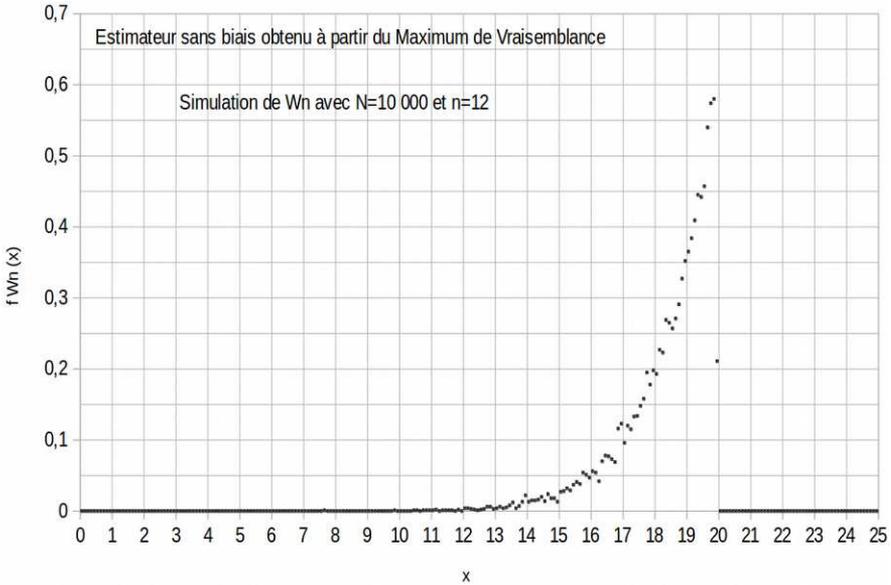
par calcul numérique avec $a \simeq 18,4$ et $n=12$,

$$\int_{a_{min}}^{20} f_{W_{12}}(x) dx = 0,9 \text{ pour } a_{min} \simeq 16,4 \text{ minutes.}$$

En conclusion $a \simeq 18,4 \begin{matrix} +1,6 \\ -2,0 \end{matrix}$ minutes à 90%.

L'intervalle trouvé est dissymétrique. Nous avons un encadrement inclus dans celui de l'inégalité de Bienaymé-Tchebychev et nous avons ici une estimation plus précise de a .

-> Effectuons maintenant une simulation numérique sur un tableur. La fonction ALEA() du tableur fournit des réels aléatoirement et uniformément répartis entre 0 et 1. La loi des $X_i = U_i(0, a)$ est alors obtenue en multipliant par l'estimation ponctuelle de a . Nous générons 10 000 échantillons de taille 12. Nous plaçons le maximum de chacun de ces échantillons sur un graphe et nous avons ainsi la distribution d'échantillonnage de T_n :



Nous trouvons ainsi les mêmes résultats que par la méthode précédente (fichier : www.incertitudes.fr/livre/Train.ods).

b) Utilisons maintenant l'estimateur du théorème des moments : $T_n = 2\bar{X}_n$. Ici, pour n grand, nous pouvons utiliser le théorème central limite, car une fonction affine d'une loi donne une nouvelle loi de même forme. Nous estimons la moyenne avec la loi de Gauss :

$$\bar{x} = \frac{3+12+\dots+11}{12} \simeq 9,83 \text{ et } s = \sqrt{\frac{(3-\bar{x})^2+\dots}{11}} \simeq 4,88$$

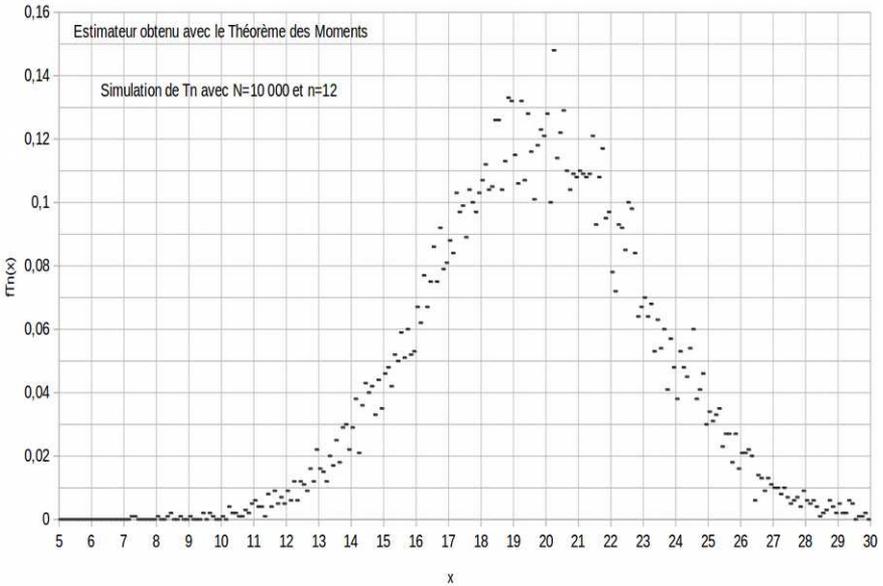
$$m = \bar{x} \pm t_{\infty} s / \sqrt{n} = 9,8 \pm 2,3 \text{ à } 90\%$$

$$\text{D'où : } a = 19,7 \pm 4,6 \text{ minutes à } 90\%.$$

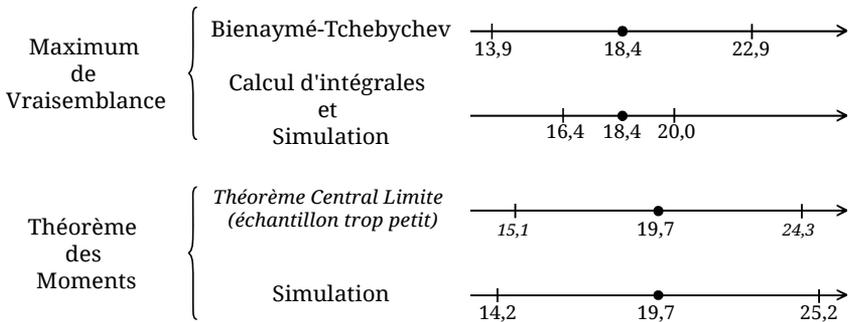
Dans ce cas n vaut seulement 12 et la loi de la population n'est pas normale, nous voulons néanmoins montrer

comment on procéderait pour n grand. Comme nous le constaterons avec la simulation numérique, l'intervalle est ainsi sous-estimé.

En effet en réalisant une simulation numérique avec cet estimateur, $a = 19,7 \pm 5,5$ minutes à 90% :

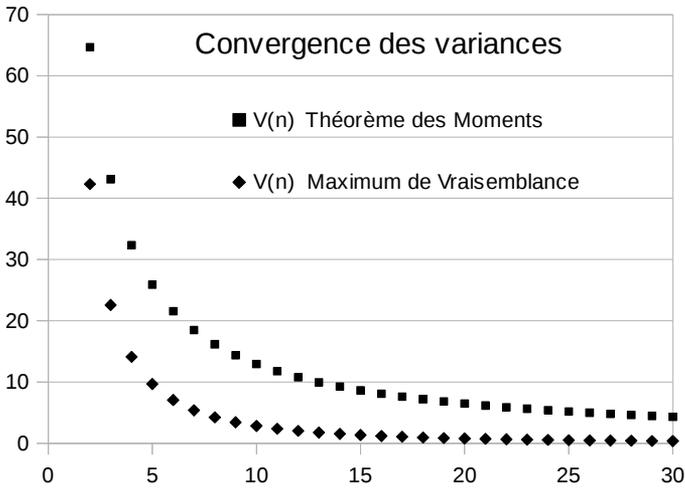


Comparaison des intervalles à 90% suivant les méthodes :



En conclusion, l'estimateur du maximum de vraisemblance converge bien plus rapidement que celui du théorème des moments et nous préférons cette première méthode. La variance converge en $1/n^2$ au lieu de $1/n$:

$$[V(W_n)]_{MV} = \frac{a^2}{n(n+2)} \quad \text{et} \quad [V(T_n)]_{TM} = \frac{a^2}{3n}$$



D. Exercices

Exercice 1 : Estimateurs de la moyenne

corrigé en version complète

Soit l'échantillon (X_1, X_2, X_3) . X_1, X_2 et X_3 sont trois variables indépendantes, de même loi, d'espérance m et de variance σ^2 .

Comparez les trois estimateurs suivants proposés pour estimer la moyenne m de l'échantillon :

$$A_3 = (X_1 + X_2 + X_3)/3, \quad B_3 = (X_1 + 2X_2 + 3X_3)/6 \text{ et} \\ C_3 = (X_1 + 2X_2 + X_3)/3.$$

Exercice 2 : Faisceau homocinétique

corrigé en version complète

Soit un faisceau homocinétique d'atomes de carbones ionisés C^+ . Nous mesurons la quantité de mouvement et l'énergie cinétique de 100 atomes du faisceau. Le faisceau est considéré parfaitement unidirectionnel.

La norme de la quantité de mouvement et l'énergie cinétique totale s'écrivent et valent :

$$p = \sum_{i=1}^{100} m v_i = 2,418 \cdot 10^{-21} \text{ kg.m/s et}$$

$$E_c = \sum_{i=1}^{100} \frac{1}{2} m v_i^2 = 1,518 \cdot 10^{-18} \text{ J avec } m = 1,993 \cdot 10^{-26} \text{ kg.}$$

Soit V la loi de probabilité de la vitesse des ions. Déterminer, grâce à l'échantillon prélevé et les estimateurs adéquats, la vitesse moyenne v_m , sa variance σ_v^2 et l'incertitude Δv à 95%.

Exercice 3 : Deux estimateurs

corrigé en version complète

Soit X la variable aléatoire discrète suivante :

Valeurs	0	1	2
Probabilités	3θ	θ	$1 - 4\theta$

1. Quelles valeurs de θ définissent une loi de probabilité ?
2. Calculer $E(X)$ et $V(X)$.
3. Soit un échantillon (X_1, X_2, \dots, X_n) de X .

Nous avons $\bar{X}_n = \frac{1}{n} \sum_{i=1}^n X_i$ et $T_n = a\bar{X}_n + b$.

Déterminer a et b pour que T_n soit un estimateur sans biais de θ .

Déterminer $V(T_n)$.

4. Soit la variable aléatoire Y_i définie pour tout $i \in \llbracket 0; n \rrbracket$ par $Y_i = 1$ si $X_i = 1$ et $Y_i = 0$ sinon.

Soit $Z_n = \sum_{i=1}^n Y_i$. Déterminer $E(Z_n)$.

Montrer que $U_n = \frac{Z_n}{n}$ est un estimateur non biaisé de

θ . Déterminer $V(U_n)$.

5. Nous effectuons des estimations de θ avec les réalisations suivantes :

Valeurs	0	1	2
Effectifs	31	12	57

Estimer θ . Quel estimateur préférez-vous ?

Exercice 4 : Urnes corrigé en version complète

Soient deux urnes identiques qui contiennent la même proportion p de boules noires. Dans la première urne nous tirons avec remise un échantillon de taille n_1 et nous notons P_1 la proportion de boules noires de cet échantillon. Nous réalisons la même expérience pour la deuxième urne.

Nous définissons $T = \frac{P_1 + P_2}{2}$ et $U = x P_1 + (1 - x) P_2$
avec $x \in]0;1[$

Montrer que T et U sont deux estimateurs de p .
Lequel est le meilleur ?

Déterminer la valeur de x permettant d'avoir l'estimateur optimum.

Exercice 5 : Variable à densité
corrigé en version complète

Soit la variable aléatoire X définie par la densité de probabilité suivante :

$$f(x) = \begin{cases} \frac{a}{x^{a+1}} & \text{si } x > 1 \\ 0 & \text{sinon} \end{cases}$$

Où a est le paramètre que nous voulons estimer ($a > 1$).

1. Montrer que f définit une loi de probabilité.
2. Calculer $E(X)$.
3. Déterminer des estimateurs de a par la méthode du maximum de vraisemblance et par la méthode des moments.
4. Donner des estimations ponctuelles de a pour les observations suivantes :
1,16 / 1,80 / 1,04 / 3,40 / 1,22 / 1,06 / 1,35 / 1,10.
5. Peut-on effectuer le calcul de $V(X)$?
6. Effectuer une simulation numérique de la loi de X .
Que pouvons nous conjecturer quant aux biais et convergences des estimateurs trouvés ?

Exercice 6 : Densité linéaire

corrigé en version complète

Soit la variable aléatoire X définie par la densité de

$$\text{probabilité suivante : } f(x) = \begin{cases} ax+b & \text{si } 0 \leq x \leq 1 \\ 0 & \text{sinon} \end{cases}$$

1. Exprimer b en fonction de a de manière à ce que f définisse une loi de probabilité.
2. Calculer $E(X)$ et $V(X)$.
3. Déterminer un estimateur T_n de a par la méthode des moments. Discuter les propriétés de cet estimateur.
4. Soient les données suivantes :

0,81	0,67	0,72	0,41	0,93	0,55	0,28	0,09	0,89
------	------	------	------	------	------	------	------	------

Déterminer une estimation ponctuelle de a . Comment pourrions-nous obtenir une estimation par intervalle ?

Exercice 7 : Estimateurs pour la loi exponentielle

corrigé en version complète

L'exercice du chapitre précédent page 124 fournit l'expression de la densité de probabilité de l'estimateur T_n de λ obtenue dans le cours :

$$f_{T_n}(x) = \frac{n^n \lambda^n}{(n-1)! x^{n+1}} e^{-\frac{n\lambda}{x}} \quad \text{si } x > 0 \text{ et zéro sinon.}$$

1. Déterminer l'espérance, le biais, la variance et le risque quadratique de T_n .
2. Soit $W_n = \frac{n-1}{n} T_n$. Déterminer l'espérance, le biais, la variance et le risque quadratique de W_n .
3. Quel estimateur conseillerez-vous pour λ ?

Exercice 8 : Désintégrations

corrigé en version complète

La loi de probabilité X de désintégration d'une particule au cours du temps suit une loi exponentielle de paramètre λ . Nous mesurons la durée de vie en microsecondes d'un échantillon de dix particules et nous voulons en déduire une estimation ponctuelle et par intervalle de λ . Nous allons utiliser différentes méthodes et les commenter.

2,20 3,32 1,93 0,88 4,36
7,10 0,65 0,53 4,69 0,31

1. En utilisant le théorème central limite, sauriez-vous estimer l'espérance m de X avec son incertitude à 90% de confiance (m : durée de vie moyenne) ? En utilisant la formule de propagation des incertitudes nous pourrions ensuite en déduire λ avec son incertitude. Que pensez-vous de cette estimation de λ ?

2. Soit T_n l'estimateur de λ trouvé en cours. Comme montré précédemment celui-ci est biaisé et nous utilisons donc $W_n = \frac{n-1}{n} T_n$.

Déterminer par un calcul d'intégrale l'incertitude à 90% sur λ .

3. Retrouver ce résultat par simulation numérique.

V. COMPLÉMENTS

A. Mesure avec une règle

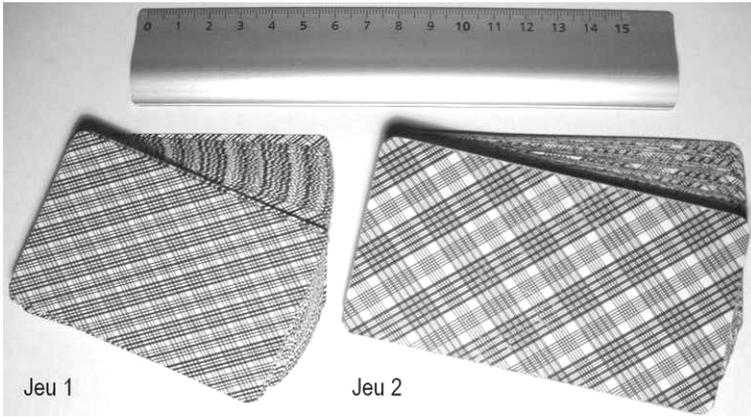
Article publié dans le BUP [ii].

RÉSUMÉ

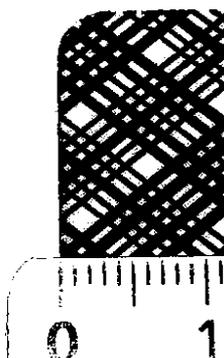
La mesure d'une grandeur par un système d'acquisition induit de part sa résolution une erreur de discrétisation. Nous nous attachons ici à la mesure d'une longueur avec une règle graduée. Ce type de mesure nous amène à considérer une loi de probabilité continue uniforme. Nous utilisons ensuite un produit de convolution pour déterminer l'incertitude avec sa confiance d'une somme de longueurs. Nous généralisons finalement au cas général du calcul d'incertitudes pour des variables aléatoires indépendantes en utilisant la formule de propagation des erreurs.

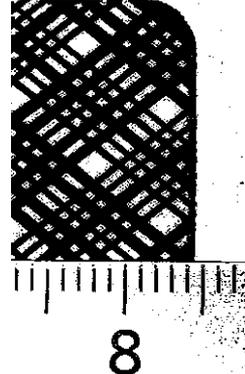
INTRODUCTION

Nous voulons mesurer des longueurs et évaluer les incertitudes le plus précisément possible. Incertitudes sur les valeurs mesurées et leurs sommes. Nous disposons d'une règle de 15cm graduée au millimètre et de deux jeux de cartes. La règle est supposée parfaite et les cartes de chaque jeu à l'identique.



1. MESURE DE LA LONGUEUR D'UNE CARTE

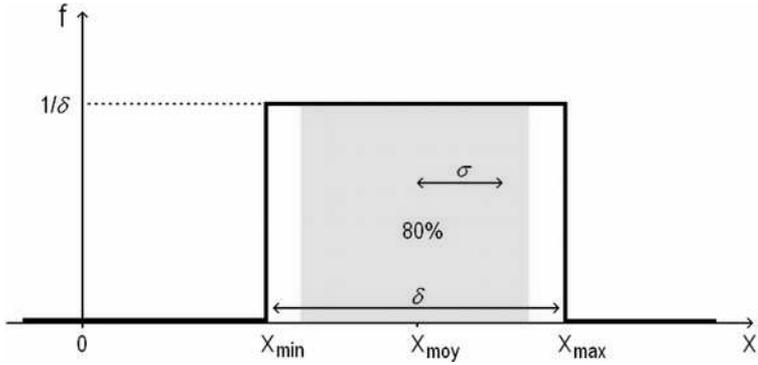

 Nous plaçons la graduation du zéro sur le bord gauche de la carte. Sur le bord droit nous considérons la graduation la plus proche du bord. L'expérimentateur ne lit pas entre les graduations. L'épaisseur des traits qui délimitent une graduation est considérée comme négligeable devant la largeur de cette graduation. Nous obtenons ainsi pour le jeu 1:



$$x_1 = 8,4 \pm 0,05 \text{ cm.}$$

Pour le jeu 2:

$$x_2 = 11,2 \pm 0,05 \text{ cm}.$$



Nous acceptons une perte d'information due à la résolution $\delta = 1 \text{ mm}$ de la règle. Lorsque ultérieurement nous exploitons ces données toutes les valeurs entre $x_{\min} = x_{\text{moy}} - \delta/2$ et $x_{\max} = x_{\text{moy}} + \delta/2$ sont équiprobables. La loi de probabilité de la variable continue aléatoire X est uniforme. x est une réalisation de X . Cette distribution de probabilité a une étendue $E = x_{\max} - x_{\min}$ et nous vérifions pour la densité de probabilité $f(x)$:

$$\int_{-\infty}^{+\infty} f(x) dx = 1$$

La probabilité pour que la valeur de X soit comprise entre x et $x + dx$ est de $f(x) dx$. Le résultat sera compris avec certitude entre x_{\min} et x_{\max} : par exemple $x_1 = 8,4 \pm 0,05 \text{ cm}$ à 100% de confiance, mais $x_1 = 8,4 \pm 0,04 \text{ cm}$ avec une probabilité de 80%.

Pour caractériser l'étalement d'une distribution considérons l'étendue E et l'écart-type σ dont la définition pour une loi continue est:

$$V = \sigma^2 = \int (x - x_{moy})^2 f(x) dx ,$$

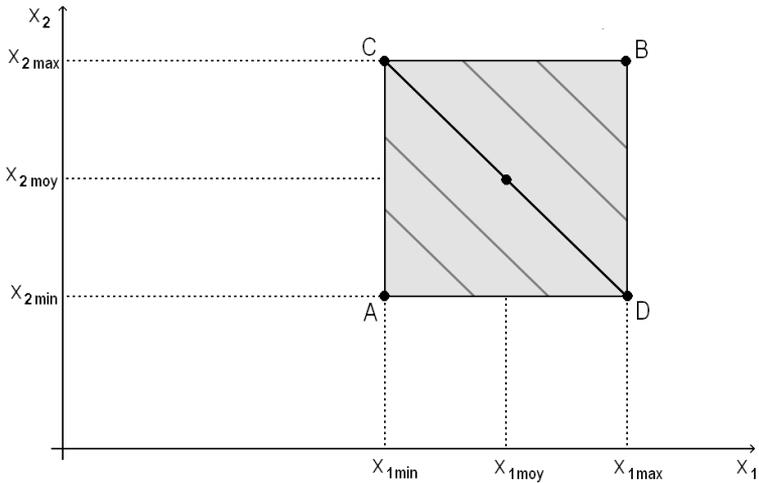
V est appelée la variance. Pour une loi uniforme:

$$\sigma = \delta / \sqrt{12} \simeq 0,29 \delta ,$$

et nous avons $x = x_{moy} \pm \sigma$ avec une confiance de 58%. L'écart-type est une grandeur adéquate pour caractériser la largeur d'une distribution. L'étendue quant à elle est définie par les valeurs extrêmes qui peuvent être peu représentatives ou pire des valeurs aberrantes.

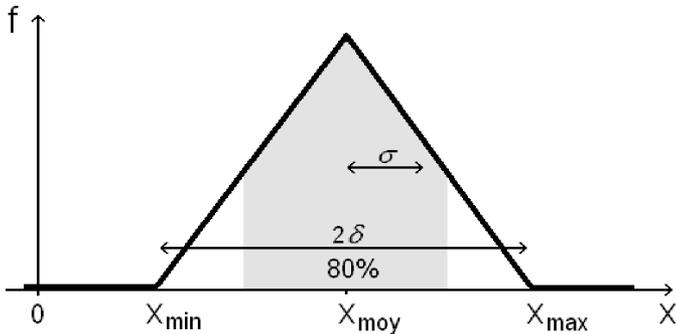
2. LONGUEUR DES DEUX CARTES MISES BOUT À BOUT

Nous souhaitons déterminer l'incertitude sur x avec $x = x_1 + x_2$. Si nous traçons x_2 en fonction de x_1 l'ensemble des points possibles forme un domaine carré. L'ensemble des points tel que x soit constant est une portion de droite de pente -1 et d'ordonnée à l'origine x : $x_2 = -x_1 + x$. Il n'y a qu'un cas qui réalise $x = x_{\min}$ soit $\left\{ x_1 = x_{1\min}; x_2 = x_{2\min} \right\}$ au point A sur la figure. Par contre sur l'ensemble du segment $[CD]$ $x = x_{moy}$. Nous comprenons que toutes les valeurs de x ne sont pas équiprobables.



La loi de probabilité f de X se calcule à partir de celle f_1 de X_1 et f_2 de X_2 . Pour une somme de variables aléatoires indépendantes le résultat est donné par un produit de convolution [iii] :

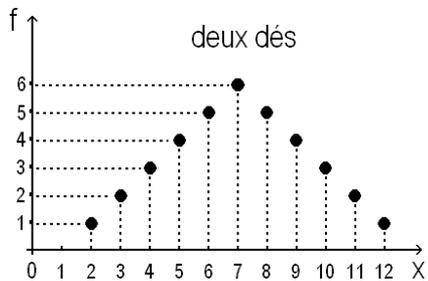
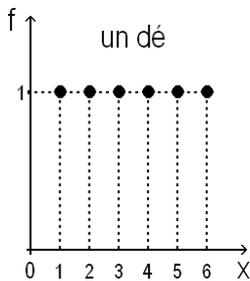
$$f(x) = \int f_1(y) f_2(x-y) dy \Rightarrow \begin{cases} x < x_{\min} \Rightarrow f(x) = 0 \\ x_{\min} < x < x_{\text{moy}} \Rightarrow f(x) = \frac{(x - x_{\min})}{\delta^2} \\ x_{\text{moy}} < x < x_{\max} \Rightarrow f(x) = \frac{(x_{\max} - x)}{\delta^2} \\ x > x_{\max} \Rightarrow f(x) = 0 \end{cases}$$



Nous avons alors une loi de probabilité triangulaire.
 Nous obtenons $x = 19,6 \pm 0,1 \text{ cm}$ avec 100% de confiance,
 et $x = 19,6 \pm 0,055 \text{ cm}$ avec 80% de confiance.

3. ANALOGIE AVEC LE LANCER DE DEUX DÉS

Pour chaque dé les six valeurs sont équiprobables. Ici la loi de probabilité n'est plus continue mais discrète. Pour le lancer simultané de deux dés, la somme des valeurs obtenues est comprise entre deux et 12. Dans ce cas il n'y a plus équiprobabilité, une manière de faire deux avec un double un, deux manières de faire trois avec un et deux ou deux et un... Pour faire sept nous obtenons le maximum de possibilités. Nous retrouvons ainsi une loi triangulaire.



4. LONGUEUR DE DEUX CARTES D'UN MÊME JEU MISES BOUT À BOUT

Les cartes d'un jeu étant supposées identiques si la longueur de l'une d'elle est surestimée, il en sera de même pour la deuxième. Dans ce cas les erreurs s'ajoutent et ne peuvent pas se compenser. Pour deux cartes différentes, la première mesure pouvait être sous-estimée et la deuxième surestimée, une compensation pouvant alors se produire. Ici ce n'est plus le cas et pour $X = X_i + X_i'$ nous obtenons une loi de probabilité à nouveau uniforme de largeur 2δ . Nos variables aléatoires ne sont plus indépendantes.

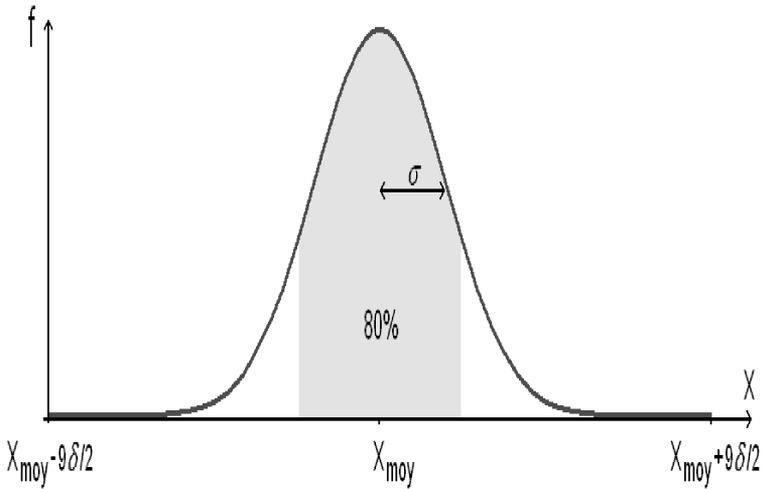
Pour le jeu 1:

$$x_1 = 8,4 \pm 0,04 \text{ cm} \Rightarrow x = 2x_1 = 16,8 \pm 0,08 \text{ cm}$$

à 80% de confiance.

5. SOMME DE N LONGUEURS INDÉPENDANTES

Nous avons $X = \sum_{i=1}^N X_i$. Chaque longueur X_i suit une loi uniforme de largeur δ . Pour la somme de neuf variables aléatoires indépendantes après itération du calcul nous obtenons la courbe suivante :



Nous avons dans ce cas $x = x_{moy} \pm 0,11 \text{ cm}$ à 80%. A 100% de confiance $x = x_{moy} \pm 0,45 \text{ cm}$ ce qui amène à considérer des domaines où la probabilité de présence de X est vraiment négligeable. Une incertitude de 0,45cm semble inutile alors que 99% des cas étaient déjà présents avec une incertitude de 0,22cm.

Raisonnement avec une confiance de 100% revient à considérer l'étendue, celle-ci est additive pour une somme de variables. L'étendue est proportionnelle à N .

	80%	95%	99%
N=1	0,40 δ	0,48 δ	0,50 δ
2	0,55 δ	0,78 δ	0,90 δ
3	0,66 δ	0,97 δ	1,19 δ
4	0,75 δ	1,12 δ	1,41 δ
5	0,84 δ	1,25 δ	1,60 δ
6	0,92 δ	1,38 δ	1,76 δ
7	0,99 δ	1,49 δ	1,91 δ
8	1,06 δ	1,59 δ	2,05 δ
9	1,12 δ	1,69 δ	2,18 δ
10	1,2 δ	1,8 δ	2,3 δ
20	1,7 δ	2,5 δ	3,3 δ
50	2,6 δ	4,0 δ	5,2 δ
100	3,7 δ	5,7 δ	7,4 δ

Mais cette approche ne tient pas compte d'une chose: la courbe se resserre autour de la moyenne quand N augmente. Il existe une autre grandeur additive: la variance. L'écart-type racine de la variance est proportionnel à \sqrt{N} et tient compte des compensations d'erreurs.

La courbe obtenue est ce qu'on appelle une courbe en cloche. Un théorème statistique, appelé théorème central limite, indique que pour N grand la courbe tend vers une gaussienne. L'étendue d'une gaussienne est infinie pour un écart-type fini.

Nous pouvons résumer l'évolution de l'incertitude sur la somme de N longueurs indépendantes mesurées avec une même résolution δ dans un tableau. En italique, à partir de N=10, il s'agit de simulations numériques réalisées sur

ordinateur par génération de nombres aléatoires.

Les résultats des mesures sont souvent donnés avec une confiance de 95%, ce qui correspond pour une gaussienne à une incertitude d'environ 2σ .

6. AUTRES APPLICATIONS

Un coureur souhaite mesurer son temps de parcours. Sa montre à affichage numérique indique qu'il part à 10 h 52 min et qu'il arrive à 11 h 11 min. L'affichage est à la minute, il est donc partagé entre 10 h 52 min 00 s et 10 h 52 min 59 s. D'où la date de départ dans l'intervalle $t_1 = 10 \text{ h } 52 \text{ min } 30 \text{ s} \pm 30 \text{ s}$. La résolution est d'une minute. La durée du parcours est $\Delta t = t_2 - t_1$. Les résultats restent vrais pour une différence. Nous avons $N=2$ et $\Delta t = 19 \text{ min} \pm 47 \text{ s}$ avec 95% de confiance.

Même démarche si des étudiants mesurent une différence d'angles sur un goniomètre. Chaque mesure étant à la minute d'arc près l'incertitude du résultat est de 47 secondes d'arc à 95%.

Sept personnes veulent rentrer en même temps dans un ascenseur. Sa charge maximale est de 500 kg. Leurs masses individuelles sont mesurées avec un pèse personne d'une résolution de un kilogramme. La masse totale est de 499 kg. Quelle est la probabilité d'être en surcharge?

Pour $N=7$ l'incertitude atteint un kilogramme avec une confiance de 80%. Il y a donc une chance sur dix pour que l'ascenseur soit en surcharge.

Au laboratoire de nombreux appareils de mesure disposent d'affichages numériques. La résolution est au dernier digit près. Mais l'incertitude globale est bien supérieure. Il faut consulter la notice de chaque appareil.

CONCLUSION

La démarche générale consiste à combiner des lois de probabilités. L'outil mathématique utilisé est un changement de variables, puis une ou plusieurs intégrations. Pour la mesure avec une règle décrite dans cet article, il s'agissait d'une somme de deux variables aléatoires indépendantes et nous avons obtenu un produit de convolution.

Si l'on veut faire un calcul plus rapide une analyse de variance peut suffire. Nous avons une variable aléatoire X qui dépend de N variables aléatoires indépendantes X_i : $X = f(X_1, X_2, \dots, X_i, \dots, X_N)$. Nous appelons σ_i l'écart-type de X_i et σ celui de X . Pour des σ_i finis et de petites variations, nous avons la formule de propagation des écart-types [iv] :

$$\sigma^2 = \sum_{i=1}^n \left(\frac{\partial f}{\partial x_i} \right)^2 \sigma_i^2$$

Et, indépendamment des lois de probabilités, cette relation entre les variances reste vraie. On pourra ainsi donner son résultat avec une incertitude à 2σ ou 3σ .

Existe-t-il une formule analogue en terme de confiance? Oui, mais elle est approximative, c'est la formule de propagation des incertitudes:

$$\Delta f^2 = \sum_{i=1}^n \left(\frac{\partial f}{\partial x_i} \right)^2 \Delta x_i^2 ,$$

avec $x_i = x_{i\text{moy}} \pm \Delta x_i$, $f = f_{\text{moy}} \pm \Delta f$ et une confiance constante. Cette formule est très pratique et permet un calcul rapide et raisonnable des incertitudes combinées. Qui plus est, elle est exacte si la forme des distributions est la même pour X et les X_i . Par exemple si les X_i sont à distribution gaussienne toute combinaison linéaire l'est aussi.

Nous tenons ainsi compte des compensations et nous évitons d'utiliser la formule $\Delta f = \sum_{i=1}^n |\partial f / \partial x_i| \Delta x_i$ qui surestime les incertitudes, parfois même avec un tel excès que l'on en perd son sens physique. Cette dernière formule ne tient compte d'aucunes compensations, on a la pire des situations, statistiquement improbable. Ici, par exemple pour $N = 100$, on aurait une incertitude de 50δ , au lieu de $5,7 \delta$ dans la pratique (confiance de 95%).

Dans cet article nous nous sommes concentré sur la résolution d'un système d'acquisition qui donne une erreur de discrétisation. Mais on peut aussi être amené à considérer des erreurs systématiques et des erreurs aléatoires. Ici la règle était supposée parfaite, c'est à dire juste et fidèle.

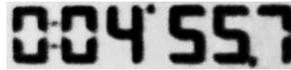
B. Métrologie

1. Exemples

Intéressons-nous à deux instruments de mesure particuliers : un chronomètre et un ampèremètre.

Mesure d'une durée avec un chronomètre :

Nous mesurons la durée d d'un phénomène. Nous supposons que l'expérimentateur appuie aux bons moments pour mesurer ce que l'on appelle les dates initiale t_1 et finale t_2 . Pour être un bon expérimentateur il faut avoir une connaissance minimale de ses instruments de mesure. Un chronomètre, du type utilisé ici, fonctionne avec un quartz très précis. La dérive est d'une quinzaine de secondes par an, soit environ 0,00004 secondes sur notre expérience. Ici, nous négligeons la dérive. Nous pouvons supposer le chronomètre parfait (juste et fidèle). Seule la résolution va intervenir ($\delta = 1/10$ éme de seconde).



$$3\text{min}32,75\text{s} < t_1 < 3\text{min}32,85\text{s}$$

$$4\text{min}55,65\text{s} < t_2 < 4\text{min}55,75\text{s}$$

$$\text{et } d = t_2 - t_1$$

Les lois de probabilité de t_1 et t_2 sont uniformes et celle de d triangulaire (voir article précédent).

D'où²² $\sigma_d = \delta/\sqrt{6} \approx 0,41\delta$, $\Delta d \approx 0,78 \delta$ (tableau page 161),

22 Calculons l'écart-type d'une distribution triangulaire (après recentrage) : $\sigma^2 = \int_{-\delta}^0 x^2 \frac{(x+\delta)}{\delta^2} dx + \int_0^{\delta} x^2 \frac{(\delta-x)}{\delta^2} dx = \frac{\delta^2}{6}$. Ou, si nous utilisons la formule de propagation des écart-types on trouve que pour une somme, ou différence, de n grandeurs de même écart-

$d = 82,900 \pm 0,078$ s et $\Delta d/d \approx 0,1\%$ avec 95% de confiance.

Maintenant, si le phénomène durait plusieurs mois, l'incertitude la plus grande serait la dérive, le chronomètre ne serait plus juste. Nous voyons que l'ensemble de la situation expérimentale est à prendre en compte et pas seulement l'instrument de mesure. Nous pourrions incorporer cette incertitude systématique de dérive par la connaissance de la valeur du biais.

Si la dérive n'est pas connue, nous pouvons prendre 100 chronomètres identiques que nous lançons simultanément. Nous aurons au final une dispersion aléatoire des valeurs qui donne une incertitude liée à la dérive.

Mesure d'un courant avec un ampèremètre :

L'ampèremètre, placé en série dans un circuit, mesure le courant de la branche où il est situé.



Nous choisissons le calibre adapté, ici pour des courants de 20 à 200 mA, et nous mesurons : $I = 186,30$ mA.

Quelle est l'incertitude sur I ?

Nous pourrions penser, comme pour le chronomètre, qu'elle est sur le dernier digit, en fait la notice indique une plus grande incertitude :

types nous avons $\sigma_n = \sqrt{n} \sigma_1$. Ici $\sigma_2 = \sqrt{2} \sigma_1 = \sqrt{2} \delta / \sqrt{12} = \delta / \sqrt{6}$.

COURANTS CONTINUS:

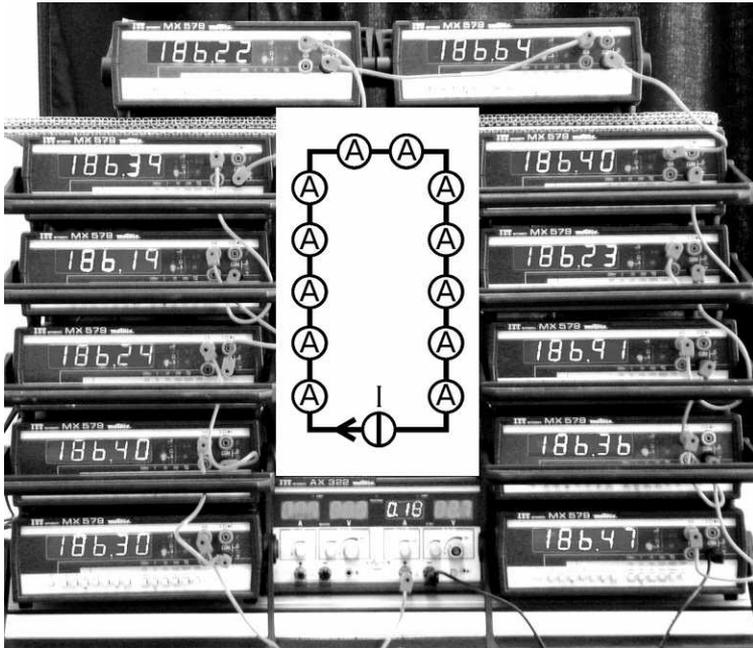
Calibre	Résolution	Précision $\pm(n\%L + mUR^*)$	Coeff. de Ture.	Chute de tension	Protection
200 μ A	10nA	0,2 % + 3 UR	2E-4/K	<300mV	fusible 2A
2mA	100nA	"	"	"	
20mA	1 μ A	"	"	"	
200mA	10 μ A	0,7 % + 3 UR	7E-4/K	"	
2A	100 μ A	"	"	<700mV	10A
10A	1mA	"	"	"	

D'où $\Delta I = 0,7\% \times 186,30\text{mA} + 3 \times 10\mu\text{A}$.

Pour simplifier nous pouvons considérer que les indications des fabricants sont données avec **une confiance de 95% et un profil gaussien.**

Alors $I = 186,30 \pm 1,34 \text{ mA}$
avec un niveau de confiance de 95%
(184,96 mA < I < 187,64 mA)

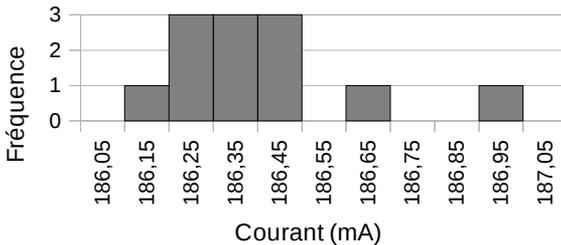
Essayons de retrouver l'erreur systématique donnée par le fabricant en regardant l'erreur aléatoire donnée par 12 ampèremètres de même modèle. Nous les plaçons en série et ainsi ils mesurent tous la même grandeur :



Le calcul de l'écart-type donne : $s_I = 0,21 \text{ mA}$, le coefficient de Student vaut $t = 2,2$, d'où $\Delta I = t \cdot s = 0,45 \text{ mA}$.

Soit $I = 186,40 \pm 0,45 \text{ mA}$ avec un niveau de confiance de 95% ($185,94 \text{ mA} < I < 186,85 \text{ mA}$).

L'incertitude est ici, à peu près, 3 fois plus faible, ce qui est normal car nous avons une meilleure connaissance de la moyenne ($\Delta \bar{I} = t \cdot s / \sqrt{n} = 0,13$



I (mA)
186,30
186,40
186,24
186,19
186,39
186,22
186,64
186,40
186,23
186,91
186,36
186,47

mA). Si nous considérons l'ampèremètre pour lequel $I=186,91\text{mA}$, seule l'incertitude systématique plus grande fournie par le constructeur garantie un bon encadrement.

Par ailleurs l'incertitude du constructeur concerne l'ensemble des ampèremètres de ce modèle qu'il produit et pas seulement ceux livrés au lycée. Aussi, nous devons considérer, pour la comparaison des résultats, que les 12 ampèremètres n'ont pas été recalibrés avec un étalon depuis leur achat d'il y a plus de 15 ans.

2. Normes

Pour déterminer rigoureusement l'incertitude, il faudrait que tous les fabricants donnent la confiance de leurs incertitudes et les profils des lois de probabilité.

Pour contribuer à une information complète sur l'expression de l'incertitude et fournir une base pour la comparaison internationale des résultats de mesure *Le comité international des poids et mesures* et *l'Organisation internationale de normalisation* ont développés un guide [vii].

Il s'agit de mettre en place des méthodes d'évaluation et d'expression des incertitudes harmonisées pour tous les pays et tous les domaines (de la santé, de la sécurité, industriel, commercial, etc.).

3. Vocabulaire

Que l'on parle avec le vocable de la statistique ou avec celui de la métrologie les notions sont les mêmes. Pour aider à faire la jonction, nous avons synthétisé certaines expressions courantes dans un tableau. Sur une même ligne vous trouverez les termes équivalents.

statistique [vi] [iv]		métrologie [v] [vii] [viii]		
grandeur à mesurer X		mesurande M		
mesure	action de mesurer		mesurage	
ensemble des mesures $\{x_i \pm \Delta x_i\}$		résultat du mesurage		
x_i	une mesure		m_i	
$X = \bar{x} \pm \Delta x$	résultat d'une mesure		$M = m \pm \Delta M$	
dispersion	fidélité (répétabilité et reproductibilité)	erreurs aléatoires (accidentelles)	erreurs de type A	
justesse	biais	erreurs systématiques	erreurs de type B	
résolution (sensibilité)	(erreur de) discrétisation			
μ (espérance)	moyenne de la population (souvent inconnue)		la valeur vraie M_{vrai}	
\bar{x}	moyenne de l'échantillon		\bar{m} ou m	
σ : écart-type de la population (souvent inconnu)				
s	s_x	écart-type de l'échantillon	écart-type expérimental	s_{exp}
s / \sqrt{n}	$s_{\bar{x}}$	écart-type estimé de la distribution d'échantillonnage	incertitude-type	$s = s_{\text{exp}} / \sqrt{n}$
coefficient de Student t $\Delta x = t \cdot s / \sqrt{n}$		facteur d'élargissement k $\Delta M = k \cdot s$		
Δx	incertitude (incertitude absolue)		incertitude-type élargie	ΔM
$\Delta x / \bar{x}$	précision (incertitude relative)		$\Delta M / m$	

Remarques :

- Plus le biais est grand, moins la mesure est juste.
- Plus les mesures sont dispersées, moins la fidélité est bonne.
- L'écart-type de l'échantillon est un estimateur biaisé de l'écart-type de la population.
- Dans ce livre nous avons évité de parler d'*erreur*, nous avons préféré le terme *incertitude*. En métrologie, nous parlons de différentes erreurs (erreur de mesure, erreur aléatoire et erreur systématique).
- Nous utilisons le terme écart pour la différence entre la moyenne de la mesure et la valeur attendue (une valeur donnée par le fabricant ou dans les tables).

4. Méthode simplifiée d'évaluation de l'incertitude de type B

Si le fabricant donne des limites et que l'on ne dispose d'aucune autre information, il est raisonnable de supposer la probabilité égale sur l'intervalle et nulle en dehors. Nous utilisons alors les caractéristiques d'une loi uniforme ([vii] p23 et [v] p7) :

Cas :	Incertitude-type :	Incertitude élargie :
Appareil à cadran, lecture d'un réglelet, digit, etc. Résolution : $\delta = x_{max} - x_{min}, x = \bar{x} \pm \delta/2$	$s_{\bar{x}} = \frac{\delta}{\sqrt{12}}$	$\Delta x = k s_{\bar{x}}$
Le constructeur fournit la classe : $\pm a$ a correspond à $\delta/2$	$s_{\bar{x}} = \frac{a}{\sqrt{3}}$	($k=2$ à 95%)

Remarques :

- Pour le premier cas nous ne disposons pas d'information du constructeur, dans le second, il fournit ce que l'on appelle la classe a de l'appareil.
- Nous prenons $k=2$ avec une confiance de 95% comme si nous avions une gaussienne (approximation).
- Si la notice est conçue selon cette norme, nous aurons pour l'ampèremètre (étudié page 167), $s_{\bar{I}} = 1,34 \text{ mA}$ et $I = 186,30 \pm 2,68 \text{ mA}$ à 95%.
- Sur le matériel de verrerie jaugé, la classe est toujours indiquée.

25 ml

Par exemple, sur une pipette jaugée de 25 mL, nous avons les indications suivantes :

**Ex 20 °C
± 0,06 ml**

La classe est 0,06 mL, d'où $s_{\bar{V}} = 0,035 \text{ mL}$ et $V = 25 \pm 0,07 \text{ mL}$ avec une confiance de 95%. La précision sur le volume est de 0,28%.

- Pour une résistance R de 1000Ω et d'une précision de 5% (brun, noir, rouge, or), nous avons :
 $\Delta R = 2 \cdot 1000 \cdot (5/100) / \sqrt{3} \approx 58\Omega$ à 95% de confiance.
 $R = 1000 \pm 58\Omega$ et $\Delta R/R \approx 5,8 \%$.
- Pesée : une balance numérique, de résolution 0,1g, indique une masse m de 73,7g.
 $\Delta m = 2 \cdot 0,1 / \sqrt{12} \approx 0,06 \text{ g}$, $m = 73,70 \pm 0,06 \text{ g}$ à 95%. $\Delta m/m \approx 0,08 \%$
- Par la méthode d'autocollimation, nous mesurons la distance focale f' d'une lentille convergente mince. L'image est nette (latitude de mise au point) de $f'_{\min} = 195 \text{ mm}$ à $f'_{\max} = 203 \text{ mm}$. D'où $s_{\text{optique}} = 2,31 \text{ mm}$.
 $s_{\text{géométrique}} = 0,29 \text{ mm}$ (δ de 1mm du banc d'optique) et $s_{\text{modélisation}} = 0,58 \text{ mm}$ (épaisseur de 2mm de la lentille).

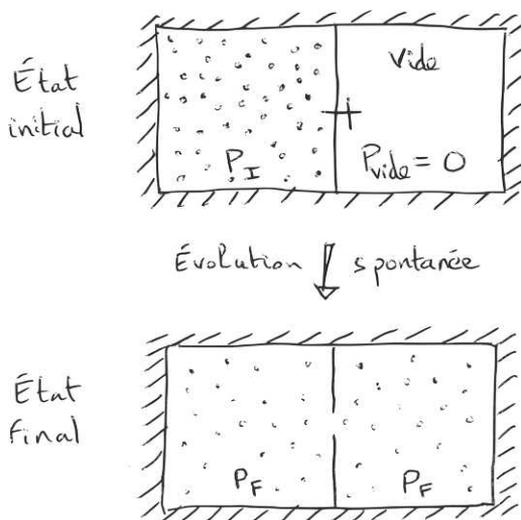
$$s = \sqrt{s_o^2 + s_g^2 + s_m^2} \quad f' = 199,0 \pm 4,8 \text{ mm} (95\%)$$

C. Thermodynamique

Nous allons étudier la détente de Joule et Gay-Lussac. Pour cette détente, nous allons utiliser deux méthodes pour déterminer les fluctuations de la pression dans l'état final : l'une classique en thermodynamique, l'autre en utilisant la formule de propagation. Dans ce cadre, nous montrerons que cette formule est en accord avec les lois de la physique.

1. Description de la détente de Joule Gay-Lussac

Nous avons deux enceintes, aux parois indéformables et calorifugées, pouvant communiquer au moyen d'un robinet. Initialement un fluide est enfermé dans l'un des compartiments, l'autre compartiment est vide. Nous ouvrons le robinet et une détente dans le vide se produit. La pression s'égalise.



Nous considérerons le cas où les deux enceintes ont le même volume.

De plus, le fluide étudié sera un gaz parfait : le premier principe de la thermodynamique permet alors de montrer que la température finale est égale à la température initiale : $T_F = T_I$; et pour les volumes $V_F = 2V_I$.

2. Hypothèses du gaz parfait

Les molécules sont considérées ponctuelles, identiques et sans interactions entre elles. Elles évoluent donc, de manière tout à fait indépendante, elles s'ignorent mutuellement, et les seuls contacts des particules sont avec les parois des récipients. Leurs chocs sur les parois exercent une pression, celle-ci dépend de la densité des particules et de la température :

$$PV = nRT \Rightarrow P = n^* k_B T$$

$$\text{avec } N = n \cdot N_A \text{ et } n^* = N/V$$

n : nombre de moles de gaz dans la boîte

N : nombre de molécules de gaz dans la boîte

n^* : densité de particules

(nombre de particules par unité de volume)

$$k_B = R/N_A : \text{constante de Boltzmann}$$

R : constante des gaz parfaits N_A : nombre d'Avogadro

Nous avons, ici, les mêmes températures initiale et finale, ainsi, $P_F = P_I/2$, et la pression est directement proportionnelle à la densité de particules :

$$P \propto n^*$$

3. Hypothèses de la physique statistique

Nous distinguons l'état macroscopique du gaz, appelé macroétat (décrit par les variables d'états de la thermodynamique, telles la pression et la température) de tous les états microscopiques correspondants, appelés microétats (décrit par les positions et les vitesses individuelles de toutes les molécules du gaz). Un expérimentateur (macroscopique) qui observe la détente ne peut mesurer que le nombre total de particules (et donc la pression) dans chaque compartiment. Il n'a pas accès à l'ensemble des informations microscopiques du système : par exemple, il ne sait pas si une particule donnée A est dans le compartiment de gauche ou de droite, il peut seulement en connaître la probabilité.

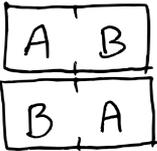
Nous noterons k le nombre total de molécules dans le compartiment de gauche.

Imaginons notre gaz constitué de cette seule particule, deux possibilités :

$$k=0 \left[\quad A \quad \right] p=1/2 \quad \text{ou} \quad k=1 \left[A \quad \right] p=1/2$$

Les deux microétats possibles ont la même probabilité p , et correspondent chacun à un macroétat.

Notre gaz a maintenant deux particules, $N=2$:

macroétats	$k=0$	$k=1$	$k=2$
microétats			
probabilités	$p=1/4$	$p=1/2$	$p=1/4$

Le macroétat $k=1$ est représenté par deux microétats.

Regardons aussi le cas $N=3$:

macroétats	$k=0$	$k=1$	$k=2$	$k=3$
microétats				
probas	$p=1/8$	$p=3/8$	$p=3/8$	$p=1/8$

Allons encore plus loin pour $N=4$:

macroétats	$k=0$	$k=1$	$k=2$	$k=3$	$k=4$
microétats	(,ABCD)	(A,BCD) (B,ACD) (C,ABD) (D,ABC)	(AB,CD) (AC,BD) (AD,BC) (BC,AD) (BD,AC) (CD,AB)	(BCD,A) (ACD,B) (ABD,C) (ABC,D)	(ABCD,)
probas	$p=1/16$	$p=1/4$	$p=3/8$	$p=1/4$	$p=1/16$

Nous voyons que, plus le nombre de particules est grand, plus les macroétats autour de $k \simeq N/2$ sont probables.

Nous avons deux principes physiques :

- les microétats sont équiprobables.
- hypothèse ergodique : au cours du temps tous les microétats sont indifféremment balayés et le système reste un temps équivalent dans chacun deux.

4. État le plus probable et fluctuations

Nous comprenons maintenant que le système sera, la plupart du temps, dans le macroétat $k=N/2$ (prenant N pair, ce qui ne restreint pas notre étude car N est très grand et nous ne sommes donc pas à une particule près).

Nombre de microétats Ω et probabilité p

pour un macroétat k donné :

$$\Omega = C_N^k = \frac{N!}{k!(N-k)!} \quad \text{et} \quad p(k) = \frac{C_N^k}{2^N}$$

Finalement, cela va se passer autour de $N/2$, alors symétrisons, et posons $k=N/2+i$, d'où : $\Omega = N!/(N/2+i)!(N/2-i)!$

Nous cherchons l'expression de $p(i)$ pour N grand et i petit, utilisons la formule de Stirling $\ln(n!) \simeq n \ln(n) - n$ avec n grand et le développement limité $\ln(1+\varepsilon) \simeq \varepsilon$ avec ε petit ; nous avons alors : $\ln[p(i)] \simeq -4i^2/N$ et $p(i) \propto e^{-4i^2/N}$.

La probabilité obtenue obéit à une loi gaussienne.

Déterminons l'écart-type : $\frac{p(\pm\sigma)}{p_{max}} = \frac{1}{\sqrt{e}}$ d'où $\sigma \propto \frac{\sqrt{N}}{2}$.

Nous avons l'incertitude relative à un sigma pour le nombre de particules et donc aussi pour la pression :

$$\frac{\sigma_{N/2}}{N/2} = \frac{1}{\sqrt{N}} \quad \text{et} \quad \frac{\sigma_P}{P} = \frac{1}{\sqrt{N}}$$

N est grand, ainsi la pression fluctue très peu et est parfaitement définie.

5. Utilisation de la formule de propagation

Nous avons N particules de pression p_j qui contribuent, ou pas, à la pression du compartiment de droite : $P = \sum_{j=1}^N p_j$.

Si la particule est dans le compartiment de gauche $p_j=0$, si elle est à droite $p_j=P_F/(N/2)$. Donc $\sigma_p = P_F/N$ (même principe que les pièces page 8). En utilisant la formule de propagation des écart-types nous obtenons : $\sigma_P = \sqrt{N} \sigma_p$

$$\text{et } \frac{\sigma_P}{P} = \frac{1}{\sqrt{N}}$$

6. Conclusion

La formule de propagation permet de retrouver rapidement le résultat de la physique statistique.

Si nous faisons l'erreur de considérer que les incertitudes s'ajoutent nous aurions $P_F = P_F \pm P_F$. Ainsi la pression et toutes les variables d'état de la thermodynamique ne seraient pas définies. De plus, cette détente, démontrée irréversible en vertu du second principe, deviendrait réversible.

Nous voyons à nouveau, sur cet exemple, que la formule de propagation n'est pas seulement un outil rigoureux pour le calcul des incertitudes mais participe, aussi, aux fondements des lois de la nature.

D. Indépendance des variables

Considérons un système. Nous mesurons deux grandeurs associées à celui-ci. Normalement les deux processus de mesure s'effectuent indépendamment. L'action de mesurer la première grandeur fournit un ensemble d'observations relatives à cette grandeur. Ces observations ont un centre et une dispersion. Dans un second temps, nous mesurons la deuxième grandeur et il n'y a aucune raison que cette deuxième collecte d'informations soit perturbée par la première. Les fluctuations de la première ne sont pas corrélées avec celles de la deuxième.

En effet, la plupart du temps, le mesurage ne modifie pas le système et les différents processus de mesure n'interagissent pas entre eux.

Exemple : nous avons un système électrique en régime stationnaire constitué d'un générateur de tension continue U qui alimente une résistance R parcourue par un courant I . Nous avons la relation théorique $U=RI$ et nous souhaitons estimer R .

Nous effectuons 100 observations de la tension U (nous pouvons prendre 100 voltmètres différents) et nous en déduisons \bar{U} et ΔU . Dans un second temps, nous procédons de même pour I , d'où \bar{I} et ΔI . Nécessairement, le coefficient de corrélation r_{UI} entre $\{U_1, \dots, U_{100}\}$ et $\{I_1, \dots, I_{100}\}$ est nul. Les grandeurs U et I sont indépendantes. D'où $\bar{R}=\bar{U}/\bar{I}$ et $\Delta R/R=\sqrt{(\Delta U/U^2+\Delta I/I^2)}$.

Deuxième exemple : nous lançons quatre dés équilibrés six faces de couleurs différentes. Nous calculons la somme : $S=d_1+d_2+d_3+d_4$. Si les lancers des trois premiers dés donnent à

chaque fois 6 cela ne préjuge en rien la probabilité d'obtenir 6 avec le quatrième dé, c'est toujours une chance sur six. Nos quatre grandeurs sont indépendantes. Conclusion : quelque soit $i, j=1,2,3,4$ avec $i \neq j$ $r_{ij}=0$ et $\sigma_s^2 = \sigma_1^2 + \sigma_2^2 + \sigma_3^2 + \sigma_4^2$.

Il existe une généralisation de la formule de propagation des écart-types dans le cas de grandeurs corrélées mais, dans la plupart des cas, nous n'avons pas à l'utiliser :

$$\sigma_f^2 = \sum_j \left(\frac{\partial f}{\partial x_j} \right)^2 \sigma_j^2 + 2 \sum_{i < j} \left(\frac{\partial f}{\partial x_i} \right) \left(\frac{\partial f}{\partial x_j} \right) r_{ij} \sigma_i \sigma_j$$

Remarques :

- $\sigma_f \leq \sum |\partial f / \partial x_i| \sigma_i$.
- si $f = x_1 + x_2$, $\sigma_1 = \sigma_2$ et $r_{12} = -1$ alors $\sigma_f = 0$.

VI. DEVOIRS

Durée de composition d'une heure pour chaque devoir. Calculatrice autorisée. Documents, ordinateurs et téléphones interdits. Toutes les réponses doivent être justifiées et clairement explicitées. Les tables de Student et de khi deux sont fournies, ainsi qu'un formulaire avec la formule de l'écart-type et celle du coefficient de corrélation.

Extraits de devoirs donnés en deuxième année d'école d'ingénieur (École Hubert Curien / Maîtrise de l'efficacité énergétique).

A. Devoir Suricate

corrigé en version complète

Exercice 1 (3 pts) :

Nous lançons quatre pièces équilibrées.

Quelle est la probabilité d'en avoir quatre qui tombent du côté pile?

Même question pour avoir trois faces et un pile suite au lancer des quatre pièces.

Nous allons maintenant nous intéresser à la composition de différentes eaux minérales :

	Calcium mg/L	Bicarbonates mg/L	Magnésium mg/L	Chlorures mg/L	Sulfates mg/L	Potassium mg/L	Sodium mg/L	pH
Evian	78	357	24	4,5	10	1	5	7,2
Badoit	190	1300	85	44	38	10	165	6
Contrex	486	403	84	10	1187	3,2	9,1	7,2
Hépar	555	403	110	11	1479	4	14	7,2
St-Yorre	90	4368	11	322	174	132	1708	6,6
Vittel	202	402	43	8	336	2	4,7	7,5
Volvic	11,5	71	8	13,5	8,1	6,2	11,6	6,95
Vichy célestin	103	2989	10	235	138	66	1172	6,8
Salvetat	253	820	11	4	25	3	7	6

Exercice 2 (7 pts) :

Nous allons tout d'abord considérer la quantité B de bicarbonates en mg/L :

- Déterminez la valeur moyenne \bar{B} de B.
- Calculez l'écart-type σ_B sur B.
- Avec une confiance de 95%, quelle est l'incertitude ΔB ?

Nous considérons maintenant la concentration S en sulfates.

- Les grandeurs B et S sont-elles corrélées ? (si la valeur absolue du coefficient de corrélation r est inférieure à la valeur critique r_c ici égale à 0,52 les grandeurs peuvent être considérées comme indépendantes).
- Déterminez la teneur moyenne totale $\{B \oplus S\}$ en mg/L des bicarbonates et sulfates avec son incertitude.

Exercice 3 (3 pts) :

Nous considérons une population d'un pays constituée de pauvres et de non-pauvres. Le pays est découpé en quatre régions. Dans chaque région nous tirons 1000 personnes au hasard et nous obtenons la répartition suivante :

Zone géographique :	Nord - Est	Centre - Nord	Sud	Ouest
Pauvres :	112	105	154	113
Non pauvres :	888	895	846	887

Nous faisons l'hypothèse que la pauvreté est uniformément répartie sur tout le pays.

L'hypothèse est-elle soutenable avec une probabilité critique de 5% ?

Exercice 4 (3 pts) :

Nous traçons maintenant pour les différentes eaux minérales la quantité de sodium en fonction de celle de chlorure. Une régression nous donne la pente et l'ordonnée à l'origine avec leurs incertitudes à 95% de confiance. Pour l'ordonnée à l'origine nous trouvons : $b = -43 \pm 25$ mg/L. Nous vous proposons des possibilités pour la pente dont une est la bonne, expliquez votre choix et pourquoi les autres ne peuvent convenir.

Pour ce faire vous devez vous aider d'une matrice de l'exercice 5²³. D'autres données qui peuvent vous aider : $M_{Na}=23,0$ g/mol, $M_{Cl}=35,5$ g/mol. Vous pourrez conclure sur

23 Devoir complet dans le livre *Probabilités, statistiques et analyses multicritères*.

l'information que l'on en tire sur les eaux minérales en général.

Cas :	1)	2)	3)	4)	5)	6)	7)	8)
pente a :	-5,34	5,34	-0,65	0,65	5,34	-5,34	0,65	-0,65
Δa :	1,19	0,19	0,19	1,19	1,19	0,19	0,19	1,19

B. Devoir Narval

corrigé en version complète

Exercice 1 (3 pts) :

Nous lançons deux dés équilibrés à six faces.

Quelle est la probabilité que la somme des deux dés fasse 12 ?

Même question pour une somme égale à 7.

Exercice 2 (7 pts) :

Pour huit maisons du même modèle nous mesurons les résistances thermiques. Pour les murs nous mesurons, de la première à la huitième maison, les valeurs R_1 suivantes en mK/W :

39	41	41	43	41	37	39	39
----	----	----	----	----	----	----	----

- a) Déterminez la valeur moyenne \bar{R}_1 de R_1 .
- b) Calculez l'écart-type σ_1 sur R_1 .
- c) Avec une confiance de 95%, quelle est l'incertitude ΔR_1 ?

Nous mesurons maintenant les résistances thermiques R_2 des huisseries. Pour les mêmes maisons, dans le même ordre, les valeurs mesurées en mK/W :

104	102	98	102	98	102	96	98
-----	-----	----	-----	----	-----	----	----

- d) Les grandeurs R_1 et R_2 sont-elles corrélées ?
- e) Déterminez la résistance thermique R de l'ensemble {murs \oplus huisseries} avec son incertitude.

Exercice 3 (3 pts) :

Nous faisons l'hypothèse d'égalité des salaires hommes-femmes au sein de l'entreprise suivante :

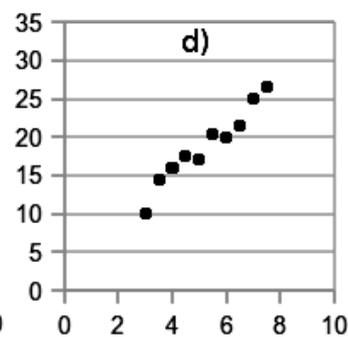
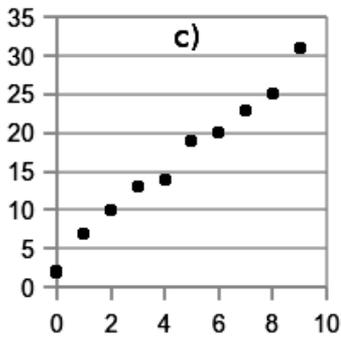
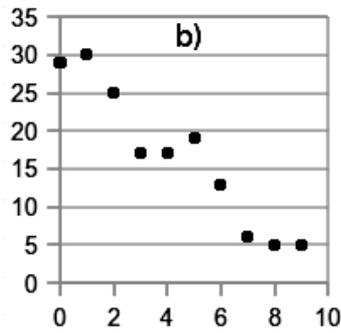
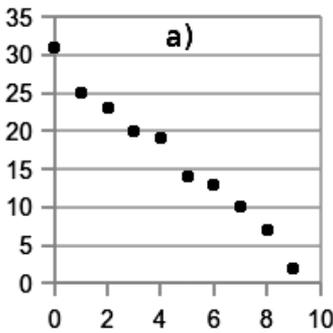
Salaires en € :	1000 à 1999	2000 à 2999	3000 à 3999	4000 à 4999
Hommes (effectifs):	139	124	68	29
Femmes (effectifs) :	261	66	22	11

L'hypothèse est-elle soutenable avec une probabilité critique de 5% ?

Exercice 4 (3 pts) :

Pour les courbes suivantes de température T ($^{\circ}\text{C}$) dans une habitation en fonction du temps t (h), indiquez si des cas du tableau correspondent à des graphes. Dans les cas positifs donnez la correspondance, sinon, expliquez pourquoi cela n'est pas possible. N'oubliez pas que vos réponses doivent être scientifiquement justifiées.

Cas :	1)	2)	3)	4)
pente a :	3,16	-5,80	2,90	-3,01
incertitude sur a :	0,63	0,31	0,31	0,69
ordonnée à l'origine b :	2,27	20,44	3,36	30,13
incertitude sur b :	3,40	1,70	1,70	3,70
coeff. de corrélation :	0,9712	-0,9914	0,9914	-0,9627



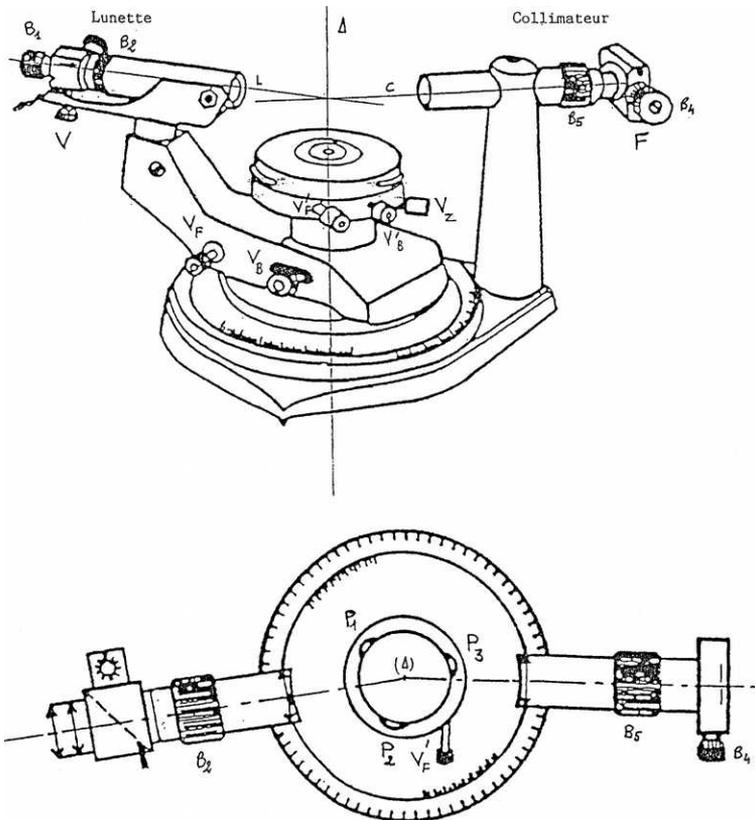
VII. TRAVAUX PRATIQUES

A. Mesure d'un indice lumineux

(TP réalisé au Lycée Alain-Fournier de Bourges)

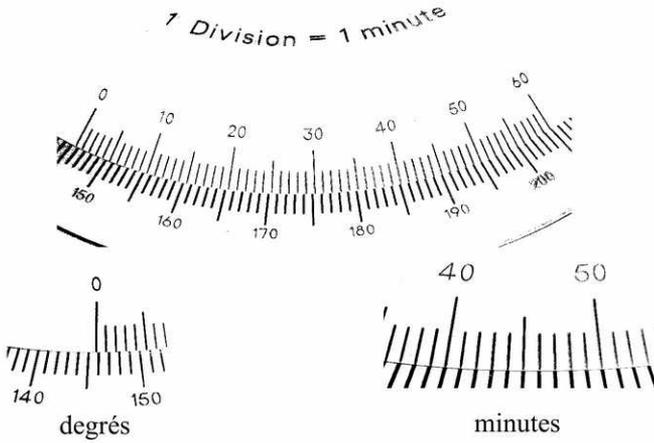
Utilisation du goniomètre et étude du prisme

Un goniomètre permet de mesurer des angles dans un plan horizontal. Il est constitué de trois parties principales mobiles autour d'un axe central vertical : le collimateur, la lunette et la plateforme.



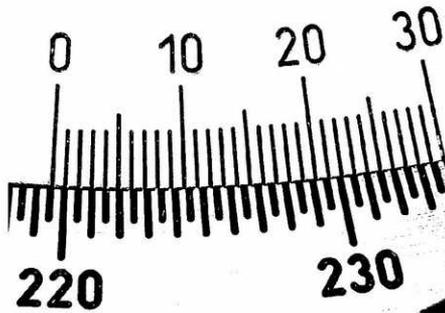
Lecture du vernier

Vernier sur 60' :



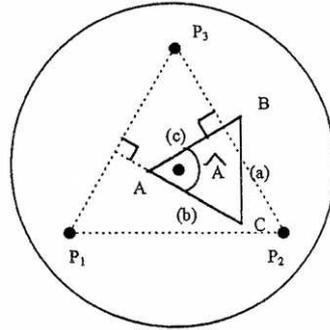
Déterminer les deux abscisses angulaires en degré et minutes (... ..)

Vernier sur 30' :



Mesure de l'angle au sommet du prisme par réflexion :

Mise en place du prisme :



Méthode par auto-collimation :

Placer la face (a) approximativement perpendiculaire à l'axe optique C du collimateur. Bloquer la plate-forme avec la vis de blocage V_a' .

Avec la lunette, viser la face B du prisme de façon à ce quelle soit perpendiculaire à (b). Déterminer avec précision la position angulaire de la lunette pour laquelle le fil vertical du réticule et son image par réflexion sur (b) coïncident : pour cela, on bloquera la rotation manuelle de la lunette avec la vis de blocage V_B au voisinage de la bonne position et on ajustera avec la vis de réglage fin V_F . Noter alors l'abscisse angulaire β_1 .

Faire de même avec la face (c). Noter β_2 correspondant.

Étude théorique :

Montrer que l'angle entre les deux positions de la lunette est égal à $\pi - A$.

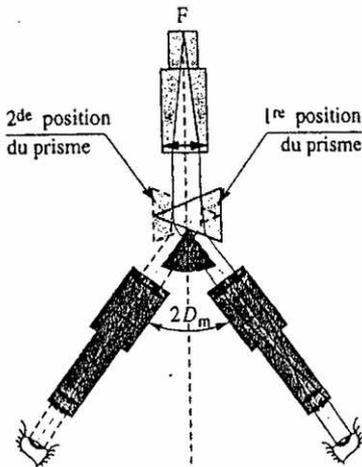
Exploitation des mesures : Dédire de ce qui précède A mesuré avec son incertitude ΔA .

Minimum de déviation et indice du prisme :

Nous plaçons la lampe à sodium derrière le collimateur. Nous obtenons un spectre de raies à la sortie du prisme. De la mesure de la déviation minimale, connaissant A, nous en déduisons l'indice via la formule du prisme donnée plus bas. Déterminez l'indice pour les raies rouge, jaune et vert-jaune. Garder le nombre de chiffres significatifs qui vous semble pertinent.

Comment évolue l'indice avec la longueur d'onde ?

Comment s'appelle ce phénomène ?



$$n = \frac{\sin\left(\frac{A + D_m}{2}\right)}{\sin\left(\frac{A}{2}\right)}$$

Les groupes suffisamment avancés peuvent mesurer les indices pour d'autres raies et tracer l'indice en fonction de l'inverse de la longueur d'onde au carré (ou mieux le logarithme de l'indice en fonction de celui de la longueur d'onde).

Conclusion ?

raies	1	2	3	4	5	6	7
	rouge	jaune	vert-jaune	vert	bleu-vert	bleu-violet	violet
$\lambda(\text{nm})$	615,7	589,3	568,5	515,2	498,1	475,0	466,7

B. Le miroir sphérique

Nous considérons deux miroirs concaves M_1 et M_2 .

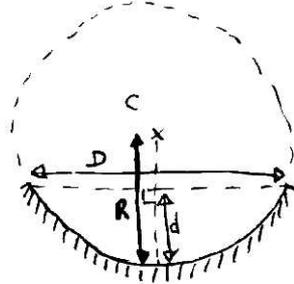
Mesure géométrique du rayon

Théorie :

Exprimez R en fonction de d et D .

Expérience : Mesurez d et D .

Estimez le rayon R avec son incertitude ΔR .



Mesure mécanique du rayon

La pulsation de petites oscillations d'une bille, sans amortissement, dans une cuvette circulaire, est $\omega_0 = \sqrt{g/R}$. Estimez R .

Mesure optique du rayon

Construction géométrique et relation de conjugaison :

Montrez que $A=A'$ avec $\gamma = -1$ pour $A=A'=C$.

Sachant cela, en déduire un protocole permettant la mesure de R .

Expérience : Mesure de $R \pm \Delta R$.

Comparez avec les autres méthodes.

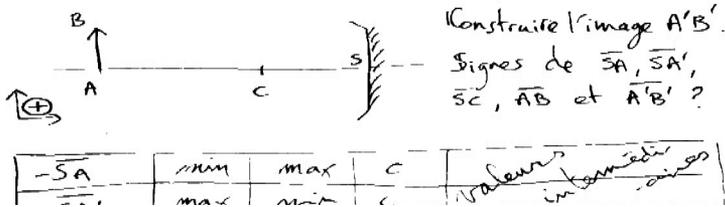
Mesure optique de la focale

On utilisera une lentille convergente telle que $A=F$ pour disposer d'un faisceau parallèle ($A'\infty$).

En déduire une méthode de mesure de la focale f . Faire la construction.

Estimez $f \pm \Delta f$, et comparez aux résultats obtenus pour R .

Relation de conjugaison



$-\overline{SA}$	(min)	(max)	(C)	(Points Inter-médiaires)
$-\overline{SA'}$	(max)	(min)	(C)	
$\Delta \overline{SA}$
$\Delta \overline{SA'}$
$x = -1 / \overline{SA}$
$y = -1 / \overline{SA'}$
Δx				
Δy				

Montrez que, d'après la théorie, nous avons une relation linéaire entre x et y . Si l'on note $y=ax+b$, à quoi doivent correspondre a et b ?

Tracer les points sur une feuille de papier millimétrée (minimum 5).

Comment estimer les incertitudes sur x et y ?

Ceci fait les rectangles d'incertitudes seront placés sur le graphique, et les droites extrêmes passant au mieux par les rectangles permettront de déterminer $a\pm\Delta a$ et $b\pm\Delta b$.

Comparaison aux valeurs attendues.

Les groupes suffisamment avancés pourront ensuite rentrer les points sur ordinateur dans le tableur prévu à cet effet et procéder par itération.

Conclusion.

C. Relation de conjugaison d'une lentille

Soit une lentille mince sphérique d'une focale de +200mm (valeur annoncée par le fabricant). Nous voulons vérifier la relation de conjugaison, et calculer la focale à partir de 8 points expérimentaux mesurés sur un banc optique de 2m. Nous plaçons successivement la lampe avec l'objet lumineux, la lentille puis l'écran (nous avons ainsi un objet et une image réels).

L'axe optique est orienté suivant la direction du rayon incident. Montrez que, d'après la théorie, nous avons une relation linéaire entre $x = -1/\overline{OA}$ et $y = 1/\overline{OA}'$. Si l'on note $y = ax + b$, à quoi doivent correspondre a et b ? Comment estimer les incertitudes sur x et y ?

Pour chaque point mesuré on prendra le plus grand soin dans l'estimation des incertitudes (incertitudes optiques dues à la latitude de mise au point, incertitudes géométriques dues à la résolution du banc, épaisseur de la lentille, etc.)

$-\overline{OA}$	(min)	(max)	(Points Intermédiaires)
\overline{OA}'	(max)	(min)	
$\Delta\overline{OA}$.	.	.
$\Delta\overline{OA}'$.		
$x = -1/\overline{OA}$.		
$y = 1/\overline{OA}'$			

Δx			
Δy			

Rentrer les points expérimentaux sur ordinateur dans le tableur prévu à cet effet et procéder par itération. Déterminez ainsi $a \pm \Delta a$ et $b \pm \Delta b$. En déduire $f' \pm \Delta f'$.
 Comparaison aux valeurs attendues.

Liste du matériel pour chaque binôme :

- un banc d'optique 2m
- 3 pieds
- une lampe source avec la flèche
- une alimentation électrique
- un écran blanc
- une lampe torche
- 1 porte lentille
- un ordinateur avec le tableur *régression avec barres d'erreurs*

Résultats obtenus par 7 groupes le 22 novembre 2007 au Lycée Alain-Fournier :

	a	Δa	f' (cm)	$\Delta f'$ (cm)
1	-0,99	0,036	19,9	0,28
2	-0,98	0,055	20,5	0,52
3	-1,07	0,036	18,9	0,29
4	-0,96	0,1	20,6	0,74
5	-1,03	0,049	20,6	0,44
6	-1,03	0,041	20,13	0,31
7	-1,02	0,059	19,58	0,42

D. Dioptries et lentilles minces sphériques

But du TP :

- Nous désirons mesurer les rayons de courbure d'une lentille mince sphérique.
- Mesurer sa distance focale par une méthode classique.
- Dédire de ces mesures l'indice de la lentille.

Théorie :

- Reprendre le théorème des vergences pour un système constitué d'une lentille et d'un miroir sphérique accolés dans les conditions de Gauss.
- Démontrez la relation de conjugaison au sommet d'un dioptré sphérique dans les conditions de Gauss. Associez deux tels dioptries accolés pour trouver l'expression de f' , focale image de la lentille, en fonction de n , R_1 et R_2 .

Proposez deux méthodes expérimentales permettant de mesurer les rayons de courbures d'une lentille mince sphérique (l'une optique, l'autre par mesure directe).

Données pour la méthode optique : Pour un rayon incident sur un dioptré, seule environ 96% de l'énergie lumineuse est transmise (réfractée), l'autre partie est réfléchi.

quez comment l'étude de cette portion réfléchié permet de remonter au rayon de courbure d'un des deux dioptre de la lentille (est-ce toujours possible ?), puis à l'indice de la lentille.

Expérience : Mesurez les rayons de courbures d'au moins deux lentilles différentes par les deux méthodes.

Proposition : celle de 200mm de grand diamètre et celle de 50cm de distance focale.

Exploitation : Une grande importance est donnée aux calculs d'incertitudes.

Mesures optiques :

	L ₁	L ₂
f _{mesuré}		
Δf _{mesuré}		
R _{mesuré}		
ΔR _{mesuré}		
R ₁ calculé		
ΔR ₁ calculé		
R' _{mesuré}		
ΔR' _{mesuré}		
R ₂ calculé		
ΔR ₂ calculé		
n calculé		
Δn calculé		

Mesure directes : Établir le même type d'exploitation.

Commentaires ? Précision sur l'indice comparée à celle obtenue avec le goniomètre ?

VIII. OUTILS MATHÉMATIQUES

Les outils mathématiques utilisés au chapitre 1 sont du niveau lycée. Les dérivées partielles utilisées au chapitre 2 sont de la première année d'enseignement supérieur mais nous comprenons très vite le lien avec les dérivées du lycée. C'est à la fin du second chapitre, avec l'utilisation des matrices pour la régression généralisée, que nous plongeons dans l'enseignement universitaire. Mon but n'est pas ici de revoir ou d'introduire toutes ces notions, seulement un petit formulaire qui pourra être utile pour résoudre les exercices proposés.

A - Dérivées

1- Définition : $f'(x) = \lim_{\epsilon \rightarrow 0} \left(\frac{f(x+\epsilon) - f(x)}{\epsilon} \right)$

Par exemple si $f(x) = x^2$:

$$(x + \epsilon)^2 - x^2 = x^2 + 2x\epsilon + \epsilon^2 - x^2 \simeq 2x\epsilon \quad \text{et} \quad f'(x) = 2x.$$

Graphiquement la dérivée correspond à la pente de la courbe en un point.

2- Principales dérivées utilisées :

fonction f	dérivée f'	Δf
ax	a (constante)	$\Delta(ax) = a \Delta x$
x^α	$\alpha x^{\alpha-1}$	$\Delta(x^\alpha) = \alpha x^{\alpha-1} \Delta x$
$\sin(x)$	$\cos(x)$	$\Delta(\sin(x)) = \cos(x) \Delta x$

$\cos(x)$	$-\sin(x)$	$\Delta(\cos(x)) = \sin(x) \Delta x$
e^x	e^x	$\Delta(e^x) = e^x \Delta x$
$\ln(x)$	$1/x$	$\Delta(\ln(x)) = \Delta x / x $
$u + v$	$u' + v'$	(u et v fonctions de x)
uv	$u'v + v'u$	
$\frac{u}{v}$	$\frac{u'v - v'u}{v^2}$	
$f(g(x))$	$g'(x) f'(g(x))$	(fonctions composées)

- $\frac{1}{x} = x^{-1}$ donc $(\frac{1}{x})' = (-1)x^{-1-1} = -\frac{1}{x^2}$.
- $\sqrt{x} = x^{\frac{1}{2}}$ donc $(\sqrt{x})' = \frac{1}{2}x^{\frac{1}{2}-1} = \frac{1}{2\sqrt{x}}$.
- $(\sin(x^2))' = (x^2)' \cos(x^2) = 2x \cos(x^2)$

B - Dérivées partielles

Celles-ci concernent les fonctions à plusieurs variables. Par exemple considérons la fonction à trois variables :

$$f(x, y, z) = x^2 - 2z + xy$$

Nous pouvons regarder les variations de cette fonction par rapport à une variable tout en considérant les autres constantes. On procède alors comme pour une dérivée. On a donc :

$$\left(\frac{\partial f}{\partial x}\right)_{y,z} = 2x + y, \quad \left(\frac{\partial f}{\partial y}\right)_{x,z} = x \quad \text{et} \quad \left(\frac{\partial f}{\partial z}\right)_{x,y} = -2$$

On parle de dérivée partielle et par exemple pour la première on dit : "dé rond" f sur "dé rond" x , à (avec) y et z constants.

C - Développement limité

Avec la notion de dérivée nous avons étudié le comportement au premier ordre d'une fonction autour d'un point, nous affinons ici aux ordres supérieurs.

Pour toute fonction infiniment dérivable et pour $\epsilon \ll 1$ nous avons le développement suivant au voisinage d'un point :

$$f(x_0 + \epsilon) = f(x_0) + \epsilon f'(x_0) + \frac{\epsilon^2}{2} f''(x_0) + \frac{\epsilon^3}{3!} f^{(3)}(x_0) + \dots + \frac{\epsilon^n}{n!} f^{(n)}(x_0) + \dots$$

Plus nous prenons de termes de haut degré plus l'approximation est précise. Par exemple pour $f(x) = \sin(x)$ et $x_0 = 0$ nous trouvons : $\sin(\epsilon) \simeq \epsilon - \frac{\epsilon^3}{3!}$, de même $\cos(\epsilon) \simeq 1 - \frac{\epsilon^2}{2}$.

$$\text{Aussi : } \exp(\epsilon) \simeq 1 + \epsilon \quad \ln(1 + \epsilon) \simeq \epsilon \quad (1 + \epsilon)^\alpha \simeq 1 + \alpha \epsilon$$

D - Intégrales

On montre en math que : $\int_{x=-\infty}^{+\infty} e^{-x^2} dx = \sqrt{\pi}$.

Pour la loi normale standard, nous pouvons vérifier que la moyenne est nulle et que l'écart-type est égal à 1 :

$$p(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} \quad \text{et} \quad \mu = \int_{-\infty}^{+\infty} x \cdot p(x) dx = 0 \quad \text{car l'intégrale sur}$$

un intervalle symétrique d'une fonction impaire est nulle.

$$\sigma^2 = \int_{-\infty}^{+\infty} x^2 \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} dx = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} (-x) (-x e^{-\frac{x^2}{2}}) dx \quad \text{d'où}$$

$$\sigma^2 \sqrt{2\pi} = [-x e^{-\frac{x^2}{2}}]_{-\infty}^{+\infty} - \int_{-\infty}^{+\infty} (-1) e^{-\frac{x^2}{2}} dx = 0 + \sqrt{2} \int_{-\infty}^{+\infty} e^{-x^2} dx$$

d'où l'écart-type égal à l'unité.

Nous avons utilisé une **intégration par parties** :

$$\int_a^b u(x)v'(x)dx = [u(x)v(x)]_a^b - \int_a^b u'(x)v(x)dx$$

puis un changement de variable $x' = x/\sqrt{2}$.

Nous pouvons aller plus loin, en calculant, le *coefficient de*

dissymétrie $\mu_3 = \int_{-\infty}^{+\infty} x^3 \cdot p(x)dx$, le *coefficient d'aplatissement*

$\mu_4 = \int_{-\infty}^{+\infty} x^4 \cdot p(x)dx$, et, de manière générale, les *moments*

d'ordre n $\mu_n = \int_{-\infty}^{+\infty} x^n \cdot p(x)dx$. L'ensemble de ces moments permettent de caractériser une distribution.

Pour une gaussienne $\mu_3 = 0$ (symétrique). Si ce coefficient est négatif la courbe s'étale vers la gauche, s'il est positif la courbe s'étale vers la droite. Pour une gaussienne $\beta_2 = \mu_4 / \sigma^4 = 3$. Si ce coefficient est inférieur à 3 la courbe est plus aplatie qu'une gaussienne. Pour une loi binomiale : $\beta_2 = 3 - 8/n$.

Changement de variable :

Soit une nouvelle variable $u = g^{-1}(x)$ avec la fonction g^{-1} continue et strictement monotone sur $[a,b]$ et g la fonction réciproque, alors :

$$\int_a^b f(x)dx = \int_{g^{-1}(a)}^{g^{-1}(b)} f(g(u))g'(u)du$$

E – Séries

Formule du binôme de Newton : $(a+b)^n = \sum_{k=0}^n \binom{n}{k} a^k b^{n-k}$

Séries géométriques dérivées : si $|q| < 1$ alors $\sum_{k=0}^{\infty} q^k = \frac{1}{1-q}$

nous dérivons ensuite par rapport à q : $\sum_{k=1}^{\infty} k q^{k-1} = \frac{1}{(1-q)^2}$

$$\text{puis } \sum_{k=2}^{\infty} k(k-1) q^{k-2} = \frac{2}{(1-q)^3} \dots$$

et finalement nous trouvons la formule du binôme négatif :

$$\sum_{k=r}^{\infty} k(k-1)\dots(k-r+1) q^{k-r} = \frac{r!}{(1-q)^{r+1}}$$

$$\text{soit } \frac{1}{(1-q)^{r+1}} = \sum_{k=r}^{\infty} \binom{k}{r} q^{k-r}$$

Une définition de la fonction exponentielle : $e^x = \sum_{k=0}^{\infty} \frac{x^k}{k!}$

F - Fonction Gamma

$$\Gamma(x) = \int_0^{+\infty} t^{x-1} e^{-t} dt$$

Cette fonction permet de généraliser la notion de factorielle à \mathbb{C} privé de \mathbb{Z}^{24} . Nous l'utiliserons pour les nombres demi-entiers.

On montre par intégration par parties que : $\Gamma(x+1) = x\Gamma(x)$

$\Gamma(1) = 1$ d'où, pour n entier, $\Gamma(n+1) = n!$.

De plus $\Gamma(1/2) = \sqrt{\pi}$ permet le calcul de la fonction pour les demi-entiers.

24 Par exemple $\pi! \simeq 7,2$.

IX. CORRECTIONS

Corrections dans le livre complet sur [lulu.com](https://www.lulu.com).

X. Bibliographie / Sources / Logiciels / Illustrations

Livres

[iv] PROTASSOV Konstantin. *Probabilités et incertitudes dans l'analyse des données expérimentales*. Presses Universitaires de Grenoble, 1999. 128 p.

[vi] WONNACOTT. *Statistique Économie - Gestion - Sciences - Médecine*. Economica, 1972. 919 p.

[x] JOURNEAUX Roger. *Traitement des mesures*. Ellipses, 2009. 377 p.

[iii] SAPORTA Gilbert. *Probabilités, analyse des données et statistique*. Technip, 2006. 622 p.

SHELDON M. ROSS. *Initiation aux probabilités*. Presses polytechniques et universitaires romandes, 2006. 585 p.

RONDY SYLVAIN. *Maths - Économique et Commerciale option Scientifique - 2^e année*. Collection Prépas Sciences, ellipse, 2014. 810 p.

ROUAUD Mathieu *Probabilités, statistiques et analyses multicritères.*, www.incertitudes.fr/proba-stat-acp/livre.pdf, 2012. 290 p, lulu.com.

Web

Placez <http://> devant le nom des sites. La plupart des fichiers sont en copie dans le dossier [<www.incertitudes.fr/livre/>](http://www.incertitudes.fr/livre/).

ROUAUD Mathieu. *Calculs d'incertitudes*. [<www.incertitudes.fr/Incertitudes.html>](http://www.incertitudes.fr/Incertitudes.html)

Calcul d'intégrales en ligne : [<www.integral-calculator.com>](http://www.integral-calculator.com)

[viii] MARCADET Françoise. *Estimer une incertitude et Mesures et incertitudes (accompagnement personnalisé en 1^{ère}S)*. Professeur de Sciences Physiques appartenant au groupe académique. Lycée Alain Fournier de Bourges. Sur le site <physique.ac-orleans-tours.fr/accompagnement_personnalise/>

[v] *Nombres, mesures et incertitudes*. mai 2010.
<www.educnet.education.fr/rnchimie/recom/mesures_incertitudes.pdf>

[vii] *Évaluation des données de mesure — Guide pour l'expression de l'incertitude de mesure*. <www.bipm.org/utills/common/documents/jcgm/JCGM_100_2008_F.pdf>, 2008. 132p.

BREUIL P, DI BENEDETTO D. *Incertitude et étalonnage*.2000. 16p.

[ix] *Algorithme de Gauss-Newton* - Wikipédia
15 juin 2013 à 17:32

Articles

[i] *Regression line analysis* K.S.KRANE et L.SCHECTER
American J. of Physics **50**,82 (1982).

[ii] ROUAUD Mathieu. *Mesure avec une règle*. Bulletin de l'Union des Physiciens n°913, avril 2009.

Logiciels

Tous les logiciels utilisés sont libres et gratuits.

- Traitement de texte et tableur *OpenOffice* et *LibreOffice*
- Graphisme : *Gimp* (points), *Inkscape* (vectoriel) et *Blender* (3D).
- Calcul : *Scilab* (numérique), *XCAS* (symbolique) et programmes en *PHP* (sur serveur).
- Tracé de courbes : *KmPlot*, *TeXgraph*.
- Système d'exploitation : *Ubuntu*.

Illustrations

Représentations 3D situées avant la table des matières et en troisième vignette de la couverture arrière.

Un marcheur aléatoire se déplace dans un plan. Il lance deux pièces et regarde les résultats : deux piles, un pile et un face, un face et un pile ou deux faces. La première pièce lui indique s'il doit faire un premier pas vers l'Est ou l'Ouest, la deuxième s'il doit faire un second pas vers le Sud ou le Nord. Ainsi à chaque intervalle de temps Δt il se déplace dans le plan de deux pas. Pendant Δt la distance parcourue par le marcheur est $\Delta d = 2 p$ (p longueur d'un pas).

À quelle distance du point de départ se trouve le marcheur à l'instant t ? (après n intervalles de temps : $t = n \Delta t$)

Pour $n=1$, nous traçons les tableaux suivants :

0,5	1	1
-0,5	1	1
	-0,5	0,5

25%	25%
25%	25%

Le centre du tableau est son point de départ. En abscisse x (direction Est-Ouest) le déplacement est de plus ou moins un pas ($x=\pm p$, $p=\Delta d/2$ et on a pris $\Delta d=1$). De même en ordonnée : $y=\pm p$. Dans chaque case est indiqué le nombre de possibilités de se retrouver à cet endroit. Le second tableau indique les probabilités ($1/4$, $4=2 \times 2$).

Dans les quatre possibilités il est à une distance $\sqrt{2}/2$ du point d'origine. Et, en terme d'écart-type, la distance caractéristique est $s=\sqrt{(4 \cdot 1/\sqrt{2})^2/(4-1)} \approx 0,816$.

Pour $n=2$, nous traçons les tableaux suivants :

1	1	2	1
0	2	4	2
-1	1	2	1
	-1	0	1

6%	13%	6%
13%	25%	13%
6%	13%	6%

Par exemple pour être en $(x=0 ; y=-1)$, il y a deux chemins possibles : (PF,PP) et (FP,PP). Écart-type $s_2 \approx 1,033$.

Le marcheur a une chance sur quatre d'être revenu au point de départ.

Pour $n=3$, nous traçons les tableaux suivants :

1,5	1	3	3	1
0,5	3	9	9	3
-0,5	3	9	9	3
-1,5	1	3	3	1
	-1,5	-0,5	0,5	1,5

2%	5%	5%	2%
5%	14%	14%	5%
5%	14%	14%	5%
2%	5%	5%	2%

$s_3 \approx 1,234$

Pour $n=4$, nous traçons les tableaux suivants :

2	1	4	6	4	1
1	4	16	24	16	4
0	6	24	36	24	6
-1	4	16	24	16	4
-2	1	4	6	4	1
	-2	-1	0	1	2

0,4%	1,6%	2,3%	1,6%	0,4%
1,6%	6,3%	9,4%	6,3%	1,6%
2,3%	9,4%	14%	9,4%	2,3%
1,6%	6,3%	9,4%	6,3%	1,6%
0,4%	1,6%	2,3%	1,6%	0,4%

Pour $n=5$, nous traçons les tableaux suivants :

2,5	1	5	10	10	5	1
1,5	5	25	50	50	25	5
0,5	10	50	100	100	50	10
-0,5	10	50	100	100	50	10
-1,5	5	25	50	50	25	5
-2,5	1	5	10	10	5	1
	-2,5	-1,5	-0,5	0,5	1,5	2,5

0,1%	0,5%	1,0%	1,0%	0,5%	0,1%
0,5%	2,4%	4,9%	4,9%	2,4%	0,5%
1,0%	4,9%	9,8%	9,8%	4,9%	1,0%
1,0%	4,9%	9,8%	9,8%	4,9%	1,0%
0,5%	2,4%	4,9%	4,9%	2,4%	0,5%
0,1%	0,5%	1,0%	1,0%	0,5%	0,1%

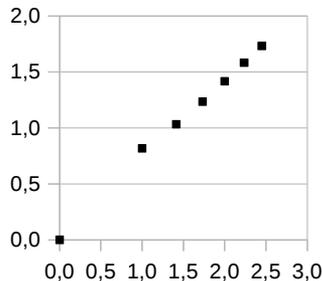
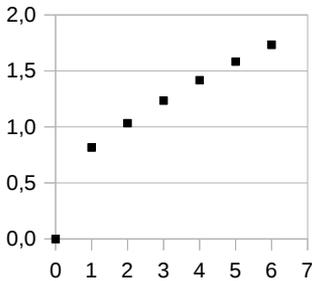
Pour $n=6$, nous traçons les tableaux suivants :

3	1	6	15	20	15	6	1	0,0%	0,1%	0,4%	0,5%	0,4%	0,1%	0,0%
2	6	36	90	120	90	36	6	0,1%	0,9%	2,2%	2,9%	2,2%	0,9%	0,1%
1	15	90	225	300	225	90	15	0,4%	2,2%	5,5%	7,3%	5,5%	2,2%	0,4%
0	20	120	300	400	300	120	20	0,5%	2,9%	7,3%	9,8%	7,3%	2,9%	0,5%
-1	15	90	225	300	225	90	15	0,4%	2,2%	5,5%	7,3%	5,5%	2,2%	0,4%
-2	6	36	90	120	90	36	6	0,1%	0,9%	2,2%	2,9%	2,2%	0,9%	0,1%
-3	1	6	15	20	15	6	1	0,0%	0,1%	0,4%	0,5%	0,4%	0,1%	0,0%
	-3	-2	-1	0	1	2	3							

D'où l'évolution de la distance quadratique moyenne du point d'origine en fonction du temps :

t (Δt)	0	1	2	3	4	5	6
\sqrt{t}	0,00	1,00	1,41	1,73	2,00	2,24	2,45
S (2p)	0,000	0,816	1,033	1,234	1,417	1,582	1,732

Nous traçons la courbe les points pour s en fonction de t , puis celle pour s en fonction de \sqrt{t} :



Nous constatons une bien meilleure corrélation en \sqrt{t} . En effet nous voyons directement que la distance à l'origine n'évoluait pas proportionnellement au temps, pour $n=2$ nous sommes à environ une unité de l'origine, nous devrions donc être vers 3 pour $n=6$.

Cette variation en \sqrt{t} est caractéristique des phénomènes de diffusion et trouve ici son analogie avec les compensations d'erreurs en \sqrt{n} .

XI. TABLES / Index

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

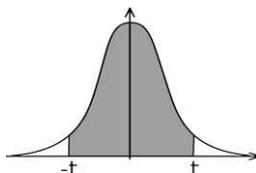
A. Loi normale centrée réduite

$$F(z) = \int_{-\infty}^z f(z) dz = P(Z \leq z) \quad P(Z > z) = 1 - P(Z \leq z)$$

$$P(Z \leq -z) = P(Z > z) \quad \text{Exemple : } P(Z \leq 1,67) \approx 0,95254$$

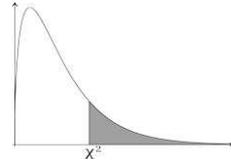
z	0,00	0,01	0,02	0,03	0,04	0,05	0,06	0,07	0,08	0,09
0,0	0,50000	0,50399	0,50798	0,51197	0,51595	0,51994	0,52392	0,52790	0,53188	0,53586
0,1	0,53983	0,54380	0,54776	0,55172	0,55567	0,55962	0,56356	0,56749	0,57142	0,57535
0,2	0,57926	0,58317	0,58706	0,59095	0,59483	0,59871	0,60257	0,60642	0,61026	0,61409
0,3	0,61791	0,62172	0,62552	0,62930	0,63307	0,63683	0,64058	0,64431	0,64803	0,65173
0,4	0,65542	0,65910	0,66276	0,66640	0,67003	0,67364	0,67724	0,68082	0,68439	0,68793
0,5	0,69146	0,69497	0,69847	0,70194	0,70540	0,70884	0,71226	0,71566	0,71904	0,72240
0,6	0,72575	0,72907	0,73237	0,73565	0,73891	0,74215	0,74537	0,74857	0,75175	0,75490
0,7	0,75804	0,76115	0,76424	0,76730	0,77035	0,77337	0,77637	0,77935	0,78230	0,78524
0,8	0,78814	0,79103	0,79389	0,79673	0,79955	0,80234	0,80511	0,80785	0,81057	0,81327
0,9	0,81594	0,81859	0,82121	0,82381	0,82639	0,82894	0,83147	0,83398	0,83646	0,83891
1,0	0,84134	0,84375	0,84614	0,84849	0,85083	0,85314	0,85543	0,85769	0,85993	0,86214
1,1	0,86433	0,86650	0,86864	0,87076	0,87286	0,87493	0,87698	0,87900	0,88100	0,88298
1,2	0,88493	0,88686	0,88877	0,89065	0,89251	0,89435	0,89617	0,89796	0,89973	0,90147
1,3	0,90320	0,90490	0,90658	0,90824	0,90988	0,91149	0,91309	0,91466	0,91621	0,91774
1,4	0,91924	0,92073	0,92220	0,92364	0,92507	0,92647	0,92785	0,92922	0,93056	0,93189
1,5	0,93319	0,93448	0,93574	0,93699	0,93822	0,93943	0,94062	0,94179	0,94295	0,94408
1,6	0,94520	0,94630	0,94738	0,94845	0,94950	0,95053	0,95154	0,95254	0,95352	0,95449
1,7	0,95543	0,95637	0,95728	0,95818	0,95907	0,95994	0,96080	0,96164	0,96246	0,96327
1,8	0,96407	0,96485	0,96562	0,96638	0,96712	0,96784	0,96856	0,96929	0,96995	0,97062
1,9	0,97128	0,97193	0,97257	0,97320	0,97381	0,97441	0,97500	0,97558	0,97615	0,97670
2,0	0,97725	0,97778	0,97831	0,97882	0,97932	0,97982	0,98030	0,98077	0,98124	0,98169
2,1	0,98214	0,98257	0,98300	0,98341	0,98382	0,98422	0,98461	0,98500	0,98537	0,98574
2,2	0,98610	0,98645	0,98679	0,98713	0,98745	0,98778	0,98809	0,98840	0,98870	0,98899
2,3	0,98928	0,98956	0,98983	0,99010	0,99036	0,99061	0,99086	0,99111	0,99134	0,99158
2,4	0,99180	0,99202	0,99224	0,99245	0,99266	0,99286	0,99305	0,99324	0,99343	0,99361
2,5	0,99379	0,99396	0,99413	0,99430	0,99446	0,99461	0,99477	0,99492	0,99506	0,99520
2,6	0,99534	0,99547	0,99560	0,99573	0,99585	0,99598	0,99609	0,99621	0,99632	0,99643
2,7	0,99653	0,99664	0,99674	0,99683	0,99693	0,99702	0,99711	0,99720	0,99728	0,99736
2,8	0,99744	0,99752	0,99760	0,99767	0,99774	0,99781	0,99788	0,99795	0,99801	0,99807
2,9	0,99813	0,99819	0,99825	0,99831	0,99836	0,99841	0,99846	0,99851	0,99856	0,99861
3,0	0,99865	0,99869	0,99874	0,99878	0,99882	0,99886	0,99889	0,99893	0,99896	0,99900
3,1	0,99903	0,99906	0,99910	0,99913	0,99916	0,99918	0,99921	0,99924	0,99926	0,99929
3,2	0,99931	0,99934	0,99936	0,99938	0,99940	0,99942	0,99944	0,99946	0,99948	0,99950
3,3	0,99952	0,99953	0,99955	0,99957	0,99958	0,99960	0,99961	0,99962	0,99964	0,99965
3,4	0,99966	0,99968	0,99969	0,99970	0,99971	0,99972	0,99973	0,99974	0,99975	0,99976
3,5	0,99977	0,99978	0,99978	0,99979	0,99980	0,99981	0,99981	0,99982	0,99983	0,99983
3,6	0,99984	0,99985	0,99985	0,99986	0,99986	0,99987	0,99987	0,99988	0,99988	0,99989
3,7	0,99989	0,99990	0,99990	0,99990	0,99991	0,99991	0,99992	0,99992	0,99992	0,99992
3,8	0,99993	0,99993	0,99993	0,99994	0,99994	0,99994	0,99994	0,99995	0,99995	0,99995
3,9	0,99995	0,99995	0,99996	0,99996	0,99996	0,99996	0,99996	0,99996	0,99997	0,99997
4,0	0,99997	0,99997	0,99997	0,99997	0,99997	0,99997	0,99998	0,99998	0,99998	0,99998

B. Coefficients de Student



Coefficient de Student t	Confiance (%)									
	50	80	90	95	98	99	99,5	99,8	99,9	
Degrés de liberté (taille de l'échantillon moins le nombre de paramètres)	1	1,00	3,08	6,31	12,7	31,8	63,7	127	318	637
	2	0,82	1,89	2,92	4,30	6,96	9,92	14,1	22,3	31,6
	3	0,76	1,64	2,35	3,18	4,54	5,84	7,45	10,2	12,9
	4	0,74	1,53	2,13	2,78	3,75	4,60	5,60	7,17	8,61
	5	0,73	1,48	2,02	2,57	3,36	4,03	4,77	5,89	6,87
	6	0,72	1,44	1,94	2,45	3,14	3,71	4,32	5,21	5,96
	7	0,71	1,41	1,89	2,36	3,00	3,50	4,03	4,79	5,41
	8	0,71	1,40	1,86	2,31	2,90	3,36	3,83	4,50	5,04
	9	0,70	1,38	1,83	2,26	2,82	3,25	3,69	4,30	4,78
	10	0,70	1,37	1,81	2,23	2,76	3,17	3,58	4,14	4,59
	11	0,70	1,36	1,80	2,20	2,72	3,11	3,50	4,02	4,44
	12	0,70	1,36	1,78	2,18	2,68	3,05	3,43	3,93	4,32
	13	0,69	1,35	1,77	2,16	2,65	3,01	3,37	3,85	4,22
	14	0,69	1,35	1,76	2,14	2,62	2,98	3,33	3,79	4,14
	15	0,69	1,34	1,75	2,13	2,60	2,95	3,29	3,73	4,07
	16	0,69	1,34	1,75	2,12	2,58	2,92	3,25	3,69	4,01
	17	0,69	1,33	1,74	2,11	2,57	2,90	3,22	3,65	3,97
	18	0,69	1,33	1,73	2,10	2,55	2,88	3,20	3,61	3,92
	19	0,69	1,33	1,73	2,09	2,54	2,86	3,17	3,58	3,88
	20	0,69	1,33	1,72	2,09	2,53	2,85	3,15	3,55	3,85
	22	0,69	1,32	1,72	2,07	2,51	2,82	3,12	3,50	3,79
	24	0,68	1,32	1,71	2,06	2,49	2,80	3,09	3,47	3,75
	26	0,68	1,31	1,71	2,06	2,48	2,78	3,07	3,43	3,71
	28	0,68	1,31	1,70	2,05	2,47	2,76	3,05	3,41	3,67
	30	0,68	1,31	1,70	2,04	2,46	2,75	3,03	3,39	3,65
	40	0,68	1,30	1,68	2,02	2,42	2,70	2,97	3,31	3,55
	50	0,68	1,30	1,68	2,01	2,40	2,68	2,94	3,26	3,50
	60	0,68	1,30	1,67	2,00	2,39	2,66	2,91	3,23	3,46
	70	0,68	1,29	1,67	1,99	2,38	2,65	2,90	3,21	3,44
	80	0,68	1,29	1,66	1,99	2,37	2,64	2,89	3,20	3,42
	90	0,68	1,29	1,66	1,99	2,37	2,63	2,88	3,18	3,40
	100	0,68	1,29	1,66	1,98	2,36	2,63	2,87	3,17	3,39
200	0,68	1,29	1,65	1,97	2,35	2,60	2,84	3,13	3,34	
300	0,68	1,28	1,65	1,97	2,34	2,59	2,83	3,12	3,32	
500	0,67	1,28	1,65	1,96	2,33	2,59	2,82	3,11	3,31	
1000	0,67	1,28	1,65	1,96	2,33	2,58	2,81	3,10	3,30	
∞	0,67	1,28	1,64	1,96	2,33	2,58	2,81	3,09	3,29	

C. Valeurs critiques de Khi-deux



χ^2	Probabilité α de rejeter l'hypothèse alors qu'elle est vraie (%)											
	99	90	75	50	30	20	10	5	2	1	0,1	
nombre de degrés de liberté	1	0,000157	0,0158	0,102	0,455	1,07	1,64	2,71	3,84	5,41	6,63	10,8
	2	0,0201	0,211	0,575	1,39	2,41	3,22	4,61	5,99	7,82	9,21	13,8
	3	0,115	0,584	1,21	2,37	3,66	4,64	6,25	7,81	9,84	11,3	16,3
	4	0,297	1,06	1,92	3,36	4,88	5,99	7,78	9,49	11,7	13,3	18,5
	5	0,554	1,61	2,67	4,35	6,06	7,29	9,24	11,1	13,4	15,1	20,5
	6	0,872	2,20	3,45	5,35	7,23	8,56	10,6	12,6	15,0	16,8	22,5
	7	1,24	2,83	4,25	6,35	8,38	9,80	12,0	14,1	16,6	18,5	24,3
	8	1,65	3,49	5,07	7,34	9,52	11,0	13,4	15,5	18,2	20,1	26,1
	9	2,09	4,17	5,90	8,34	10,7	12,2	14,7	16,9	19,7	21,7	27,9
	10	2,56	4,87	6,74	9,34	11,8	13,4	16,0	18,3	21,2	23,2	29,6
	11	3,05	5,58	7,58	10,3	12,9	14,6	17,3	19,7	22,6	24,7	31,3
	12	3,57	6,30	8,44	11,3	14,0	15,8	18,5	21,0	24,1	26,2	32,9
	13	4,11	7,04	9,30	12,3	15,1	17,0	19,8	22,4	25,5	27,7	34,5
	14	4,66	7,79	10,2	13,3	16,2	18,2	21,1	23,7	26,9	29,1	36,1
	15	5,23	8,55	11,0	14,3	17,3	19,3	22,3	25,0	28,3	30,6	37,7
	16	5,81	9,31	11,9	15,3	18,4	20,5	23,5	26,3	29,6	32,0	39,3
	17	6,41	10,1	12,8	16,3	19,5	21,6	24,8	27,6	31,0	33,4	40,8
	18	7,01	10,9	13,7	17,3	20,6	22,8	26,0	28,9	32,3	34,8	42,3
	19	7,63	11,7	14,6	18,3	21,7	23,9	27,2	30,1	33,7	36,2	43,8
	20	8,26	12,4	15,5	19,3	22,8	25,0	28,4	31,4	35,0	37,6	45,3
	21	8,90	13,2	16,3	20,3	23,9	26,2	29,6	32,7	36,3	38,9	46,8
	22	9,54	14,0	17,2	21,3	24,9	27,3	30,8	33,9	37,7	40,3	48,3
	23	10,2	14,8	18,1	22,3	26,0	28,4	32,0	35,2	39,0	41,6	49,7
	24	10,9	15,7	19,0	23,3	27,1	29,6	33,2	36,4	40,3	43,0	51,2
	25	11,5	16,5	19,9	24,3	28,2	30,7	34,4	37,7	41,6	44,3	52,6
	26	12,2	17,3	20,8	25,3	29,2	31,8	35,6	38,9	42,9	45,6	54,1
	27	12,9	18,1	21,7	26,3	30,3	32,9	36,7	40,1	44,1	47,0	55,5
	28	13,6	18,9	22,7	27,3	31,4	34,0	37,9	41,3	45,4	48,3	56,9
	29	14,3	19,8	23,6	28,3	32,5	35,1	39,1	42,6	46,7	49,6	58,3
	30	15,0	20,6	24,5	29,3	33,5	36,3	40,3	43,8	48,0	50,9	59,7
	31	15,7	21,4	25,4	30,3	34,6	37,4	41,4	45,0	49,2	52,2	61,1
	32	16,4	22,3	26,3	31,3	35,7	38,5	42,6	46,2	50,5	53,5	62,5
	33	17,1	23,1	27,2	32,3	36,7	39,6	43,7	47,4	51,7	54,8	63,9
	34	17,8	24,0	28,1	33,3	37,8	40,7	44,9	48,6	53,0	56,1	65,2
	35	18,5	24,8	29,1	34,3	38,9	41,8	46,1	49,8	54,2	57,3	66,6
	36	19,2	25,6	30,0	35,3	39,9	42,9	47,2	51,0	55,5	58,6	68,0
	37	20,0	26,5	30,9	36,3	41,0	44,0	48,4	52,2	56,7	59,9	69,3
	38	20,7	27,3	31,8	37,3	42,0	45,1	49,5	53,4	58,0	61,2	70,7
	39	21,4	28,2	32,7	38,3	43,1	46,2	50,7	54,6	59,2	62,4	72,1
	40	22,2	29,1	33,7	39,3	44,2	47,3	51,8	55,8	60,4	63,7	73,4

Index

- Asymptotes.....96
Biais.....127
Calcul d'incertitude.....54
Capacité thermique.....1
Changement de variable. 201
Chi 2.....77
Classe.....6, 38
Clous.....100
Coefficient d'aplatissement
.....103, 201
Coefficient de corrélation.48
Coefficient de dissymétrie
.....103, 201
Coefficient de Student**12**, 61,
210
Coefficients binomiaux...104
Compensations.....55, 161
Conductivité thermique....91
Confiance.....12
Décomposition en
gaussiennes.....100
Densité de probabilité.....**19**,
155
Densité linéaire.....150
Dérivée partielle.....60, 199
Dés.....16, 38, 158
Désintégrations.....151
Détente de Joule Gay-Lussac
.....173
Développement limité.....82,
200
Développement limité,....107
Diffusion.....208
Dispersion.....170
Distribution
d'échantillonnage.....10, 24
Distribution de Gauss. 10, **19**
Droites extrêmes.....62
Droites extrêmes.....193
Eaux minérales.....181
Écart.....171
Écart moyen.....3, 37
Écart quadratique moyen...4
Écart-type.....3, 170
Équation différentielle.....92
Erreur accidentelle....33, 170
Erreur aléatoire.....164, 170,
171
Erreur de discrétisation...33,
153, 170
Erreur de mesure.....171
Erreur de type A.....170
Erreur de type B.....170
Erreur systématique 33, 164,
170, 171
Espérance.....20, 170
Estimateur.....127

Estimation.....	62	Incertitude.....	13
Estimation par intervalle	127,	Incertitude absolue.....	13 , 56
138		Incertitude relative.....	13 , 56
Estimation ponctuelle.....	127	Incertitude-type.....	170
Étendue.....	3, 155	Indice.....	89, 187
Étoiles filantes.....	108	Inégalité de Bienaymé- Tchebychev.....	138, 141
Facteur d'élargissement...	170	Intégrale.....	20, 200
Faisceau homocinétique..	147	Intégrale multiple.....	42
Fidélité.....	33	Intervalle de confiance....	12,
Fluctuations statistiques....	12	62, 96	
Focométrie.....	89, 94, 172	Intervalle de fluctuation....	12
Fonction de répartition. 106,		Intervalle de prédiction...	63,
115		96	
Fonction Gamma.....	202	Intervalle dissymétrique..	143
Fonction réciproque.....	115	Isolant.....	39
Fonctions de variables à		Justesse.....	33, 170
densité.....	115	Linéarisation.....	68, 81
Formule de Cauchy.....	78, 90	Loi binomiale. 104 , 120, 121	
Formule de propagation des		Loi binomiale négative..	107,
écart-types.....	53, 163	121	
Formule de propagation des		Loi de Bernouilli.....	14
incertitudes.....	53, 164	Loi de Bernoulli.....	120
Formule du binôme de		Loi de durée de vie. 106, 111	
Newton.....	202	Loi de l'inverse.....	124
Formule du binôme négatif		Loi de Poisson 107, 122, 131	
.....	202	Loi de probabilité	
Fréquence.....	6	triangulaire.....	110, 158, 165
Gaussienne 3D.....	42	Loi de probabilité uniforme	
Gaz parfait.....	66, 174	109, 155
Goniomètre.....	187	Loi de Student.....	112, 122
Hypothèse ergodique.....	176		

Loi du Khi-Deux.....	113	Moyenne arithmétique.....	2
Loi du produit.....	123	Moyenne géométrique.....	2
Loi exponentielle. .	111, 122, 132	Nuages de points.....	50
Loi géométrique.....	105 , 121	Ordonnée à l'origine.....	60
Loi normale.....	112, 133	Parfaite.....	164
Loi uniforme.....	122, 133	Pente.....	60
Lois continues.....	109	Physique statistique.....	175
Lois discrètes.....	104	Pile électrique.....	94
Lois du Khi-Deux.....	123	Prédiction.....	63
Macroétat.....	175	Prisme.....	90, 187
Médiane.....	2	Produit de convolution. .	153, 157
Mesurage.....	170	Proposition contraposée....	25
Mesurande.....	170	Puissance du test.....	29
Mesure optique.....	192	Régression avec barres d'erreurs.....	65
Méthode d'inversion.....	118	Régression généralisée.....	76
Méthode de Box-Müller. .	119	Régression linéaire.....	59
Méthode des moments....	130	Régression multiple.....	76
Méthode des petites variations.....	74	Régression non linéaire....	81
Méthode du maximum de vraisemblance.....	134	Régression parabolique....	79
Méthode itérative. .	65, 78, 81	Régression polynomiale....	78
Métrologie.....	165	Répétabilité.....	33, 170
Microétat.....	175	Reproductibilité.....	33, 170
Miroir sphérique.....	191	Résidu.....	60, 77
Mode.....	1	Résistance thermique	91, 184
Moindres carrés....	59, 65, 76	Résolution.....	33, 153, 170
Moment.....	103, 201	Risque de deuxième espèce	26
Moyenne.....	2	Risque de première espèce	

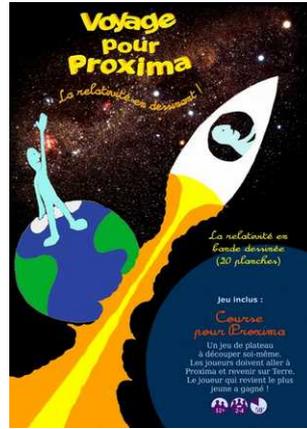
.....	26	Temps d'attente.....	139
Risque quadratique.....	128	Test d'hypothèse.....	24
Séries géométriques dérivées		Test du Khi-deux.....	30
.....	202	Théorème central limite..	10 ,
Simulation numérique...118,		24, 139, 161	
143		Train.....	139
Somme d'exponentielles. 124		Urnes.....	149
Somme de binomiales.....	121	Valeur vraie.....	170
Somme de gaussiennes...122		Variance...20, 103, 156, 161	
Somme de Students.....	123	Vernier.....	188
Somme de variables		Vieillessement.....	106, 111
aléatoires.....	103	Vraisemblance.....	134
Spectre de raie.....	90	Zéro absolu.....	63

Du même auteur :

Voyage pour Proxima *La relativité en dessinant !*

Un livre pour tous les curieux de science. Ouvrage pédagogique et abondamment illustré. Les secrets de la dilatation du temps et de la structure de l'espace sont mis en lumière de manière originale, complète et sans la moindre équation. Ce livre à la main, partez à l'aventure et rejoignez les étoiles !

Ouvrages disponibles sur lulu.com



Probabilités, statistiques et analyses multicritères

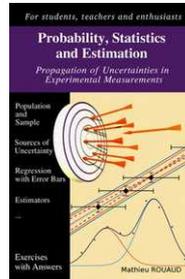
Livre qui reprend le contenu de *Calcul d'incertitudes* avec en plus les approches multicritères comme l'analyse en composantes principales.



Probability, Statistics and Estimation

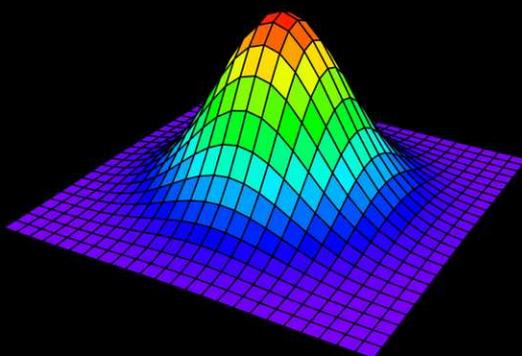
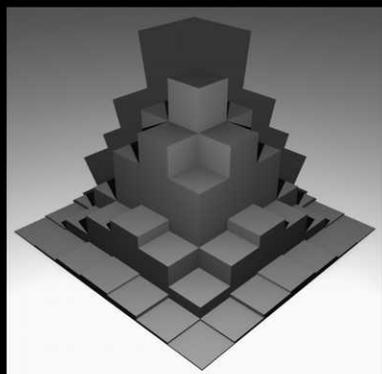
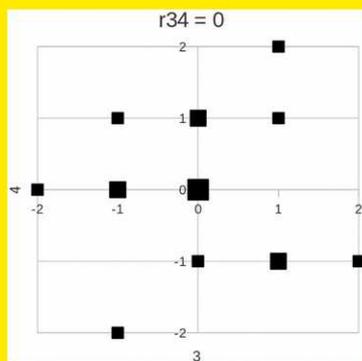
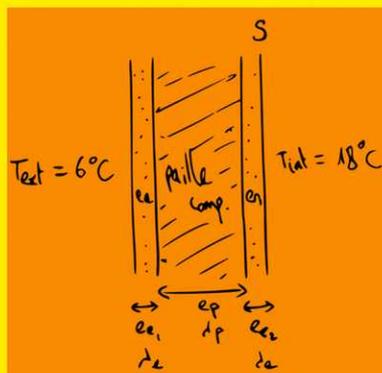
Propagation of Uncertainties in Experimental Measurement

La traduction en anglais de *Calcul d'incertitudes*.



Ce livre s'adresse à tout public intéressé par les sciences expérimentales et les mathématiques. Les statistiques et les sondages sont très présents dans notre société et le champ d'application se veut le plus large possible. Nous souhaitons dépasser l'approche purement théorique pour que l'expérimentateur trouve tous les outils nécessaires à des calculs d'incertitudes simples et rigoureux.

En effet, la science s'efforce de lier les phénomènes naturels à une logique mathématique. Cette quête d'une vérité et d'une cohérence globale nécessite d'exercer un esprit critique qui s'appuie sur des mesures accompagnées de leurs incertitudes.



ISBN 978-2-9549309-0-9



CC creative commons